Mechanisms of chaperonin-assisted protein folding

MECHANISMS OF CHAPERONIN-ASSISTED PROTEIN FOLDING

Thesis submitted to **Birkbeck, University of London**,
for the degree of **Doctor of Philosophy**, 2021

Scott Gardner

Birkbeck, University of London

Declaration of originality

CryoEM grid preparation using the Chameleon instrument was performed by Michele Darrow (SPT Labtech). Collection of high resolution cryoEM data was performed with assistance from Natasha Lukyanova. Native mass spectrometry was performed with assistance from Charles Eldrid and Hannah Britt.

The work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the text.

Scott Gardner, 2021

Abstract

The bacterial chaperonin GroEL-GroES promotes protein folding through ATP-regulated cycles of substrate protein binding, encapsulation, and release. Here, we have used recent advances in cryoEM to determine cryoEM structures of GroEL, GroEL-ADP·BeF_x, and GroEL-GroES-ADP·AlF_x, all bound to the model substrate *Rhodospirillum rubrum* RuBisCO in a non-native state. The use of the Chameleon vitrification instrument reduced preferred orientation and air-water interface denaturation effects in cryoEM. Previous cryoEM reconstructions of GroEL-substrate complexes have been limited in resolution, and there are no previously reported structures of GroEL-ATP (or ATP analogue) bound to a non-native substrate. Our structures show how the conformation and interactions of a non-native substrate protein change as GroEL proceeds through its reaction cycle. We observe specific charged and hydrophobic GroEL residues forming strong initial contacts with non-native RuBisCO. Binding of an ATP analogue results in an enhanced interaction of substrate with the disordered C-terminal tails of GroEL. The cryoEM structure of GroEL-ADP·BeF_x-RuBisCO displays striking asymmetry in the substrate-occupied ring. We observe individual GroEL subunits fulfilling different functions. Four of the GroEL subunits sequester non-native substrate inside the cavity while the remaining three adopt an extended conformation allowing GroEL to efficiently recruit GroES. Our structure explains how GroEL can recruit GroES without releasing bound substrate. The cryoEM structure of GroEL-GroES-RuBisCO shows a native-like RuBisCO interacting weakly with the walls of GroEL and the inner face of GroES. Additionally, the *trans* ring of our nucleotide bound structures show a novel conformation that may be related to the negative cooperativity of non-native RuBisCO-binding.

Table of Contents

Acknowledgements

First and foremost, I want to thank my supervisor Helen Saibil for her patience, guidance, and support. I am grateful for everything I have learned over the last four years, from cryoEM theory to the finer points of academic grammar. I also wish to thank my secondary supervisor Kostas Thalassinos for his enthusiasm and support over the years.

Nobody finishes a PhD by themselves. The staff at the ISMB have been instrumental in my success. Thank you to Claire Bagneris for running the Franklin lab and for assistance running native PAGE experiments, and helping out when I was new to purifications, to Dave Houldershaw for keeping the servers running and troubleshooting any kind of problem in five minutes or less, to Natasha Lukyanova and Shu Chen for running the EM lab, collecting cryoEM data and for all the training, advice, and friendly chatter over the years.

For the protocols and advice on GroEL, GroES and Rubisco purifications, I would like to thank Dr. Wayne Fenton, who also supplied me with the Rubisco expression plasmid. Pete Lund, who I met at the UK chaperone club one year, was more than happy to send over GroEL-GroES expression plasmids. For advice on sample preparation and for help with Chameleon grid preparation at eBIC (which sadly did not make it into this thesis), I thank Daniel Clare. The cryoEM reconstructions in this work would not be as good as they are without the assistance of Michele Darrow, who operated the Chameleon instrument for grid preparation.

My cryoEM training was possible thanks to help from past group members Joe Beton, Celia Deville and Natasha Burgess, as well as the EM lab managers Natasha and Shu. I have had productive and fun discussions with all members of Helen's lab, past and present. Everyone made Birkbeck a fantastic place to complete a PhD and I am forever grateful for the opportunity and experience.

Outside of academia, I first thank my decade-long girlfriend Alex, who has helped me through everything with unwavering support that cannot adequately be put into words. Thank you to my friends Rolf, Levi and Oleg, for providing much needed gaming downtime over the years. Cheers to my brother Glen, who constantly referred to this 200-page thesis as my "5-word essay", and cheers to my stepdad Steve, who kept me topped up with Scotch and is a constant source of (bad) humour. Lastly, I want to thank my mum, to who I dedicate this PhD thesis. Thank you for your inspiration, motivation, and help over the years!

Chapter I

1. Introduction

In cells proteins are synthesised as linear chains of amino acids. They account for around half of the total dry mass of cells and provide organisms with a vast array of functionality. To carry out their biological functions, most proteins must fold to form specific three-dimensional structures. A correctly folded protein is said to have adopted its native state. Some human diseases are related to the accumulation of toxic, non-native states of proteins in cells. How proteins adopt their native states and how cells maintain proteostasis has been an ongoing topic of biological and clinical relevance for sixty years.

1.1. Protein folding

1.1.1. Early work

Pioneering experiments performed by Christian Anfinsen and colleagues in the 1960s provided an early insight into protein folding. They showed that unfolded ribonuclease could recover its enzymatic activity following removal of the denaturant. This prompted Anfinsen to hypothesise that all the information required for a protein to adopt its native state was contained solely in the amino acid sequence (Anfinsen et al., 1961). Further work on ribonuclease conducted throughout the 1960s led to Anfinsen receiving the 1972 Nobel Prize in Chemistry (Nobel Media AB, 2021). He and others provided evidence for Anfinsen's thermodynamic hypothesis, which proposed that the native state of a protein was the state with the lowest Gibbs free energy and therefore the most thermodynamically stable (Anfinsen, 1973).

Cyrus Levinthal posited that due to the large number of states that a polypeptide chain can adopt, proteins cannot reach their native states simply by sampling conformations randomly. He thus

speculated on the existence of folding pathways (Levinthal, 1968). For a protein consisting of 2,000 atoms (~100 amino acids), there would be 10^{300} possible configurations (Levinthal, 1969). Considering that isomerisation rates for protein chains are on the order of 10^{-12} s, it would take many times the age of the universe for a protein to attain its native state through unbiased random sampling (Plotkin and Onuchic, 2002). It was clear to Levinthal and others that protein folding could not be a random search problem, as proteins were known to fold rapidly in the cell. Later work measured the folding times of many proteins, setting a range between microseconds (Qiu et al., 2002) and minutes (Goldberg et al., 1990). The discrepancy between the time scale required for an exhaustive search and the rapid protein folding times seen *in vivo* became known as the Levinthal paradox (Karplus, 1997). It has formed the basis of many models of protein folding (Englander and Mayne, 2017).

1.1.2. Early models of protein folding

Many early models describing the kinetics of protein folding were proposed. The two main models were i) the nucleation-propagation (or sequential) model (Wetlaufer, 1973) and ii) the diffusion-collision model (Karplus and Weaver, 1976). Both attempted to resolve the Levinthal paradox by explaining how the conformational space sampled by a folding protein was restricted.

Briefly, the nucleation-propagation model supposed that small stretches of amino acids would sample conformations randomly until they folded to their native state. These folded regions would act as nucleation sites and very rapidly propagate folding through the remainder of the amino acid chain (Wetlaufer, 1973). The second model, the diffusion-collision model, viewed protein structures as a collection of several structural microdomains. Each microdomain was short enough to rapidly (and randomly) reach its native state. Several structural domains would then diffuse together and coalesce to form the global native state (Karplus and Weaver, 1976). Although the models are not entirely mutually exclusive, the rapid nucleation-propagation mechanism generally precluded the population of intermediate states. Thus, it fell out of favour when it was

demonstrated that folding of some proteins did include populated intermediate states (Kim and Baldwin, 1990).

Alongside the development of the main kinetic models, other aspects of protein folding mechanisms were investigated. Proline isomerisation was identified as an important rate-limiting step in the folding of some proteins. This was discovered after RNase A was shown to exhibit two-state kinetics (Brandts et al., 1975). More recent work has also uncovered further intricacies in protein folding, such as the presence of knots in the native states of 1% of proteins, though the functional relevance of these is still unclear (Soler and Faísca, 2013).

1.1.3. The 'new view' of protein folding

During the 1990s a new view of protein folding emerged based on the general characteristics of the energy surfaces of proteins (Baldwin, 1994 and Karplus, 2011). This view was derived from statistical mechanical modelling and early computer simulations (Bryngelson and Wolynes, 1987 and Sali et al., 1994).

Significant insight into the nature of folding free-energy landscapes came from lattice simulations, such as those by Sali et al. (1994). In these simulations, small, simplified polypeptides were simulated on a 3x3x3 cubic lattice and subjected to a random search for the global native state. Starting from 10^{16} random coil conformations, the model protein underwent a fast collapse to 10^{10} semi-compact states followed by a rate-limiting step to 10^{3} native-like transition states. From one of these transition states the chain rapidly reached the native state. Later molecular dynamics simulations showed results in agreement with this model (Cavalli et al., 2003).

The key finding of these simulations was the presence of many different folding pathways to the native state rather than a single pathway. The work also showed that the native state of a fast-folding sequence was at a deeper global energetic minimum than those of slow or non-folding sequences. This latter observation led the authors to suggest an evolutionary mechanism in which

slow or non-folding sequences were more likely to be degraded in the cell, leading to positive selection of fast-folding sequences. The simulations demonstrated the first concrete solution to the Levinthal paradox (Karplus, 2011) and contributed to the development of the energy landscape theory of protein folding (Wolynes et al., 1995). A common way to represent this solution to the Levinthal paradox is with a folding funnel diagram (**Fig. 1-1**).

Computer simulations of protein folding have since grown in accuracy and complexity. Early lattice simulations used highly simplified models of proteins and chiefly modelled intramolecular interactions. Molecular dynamics (MD) simulations offer the ability to compute all-atom simulations of proteins and solvent over increasingly longer time scales. MD simulations have become an important tool in protein folding and dynamics research (Lindorff-Larsen et al., 2011).

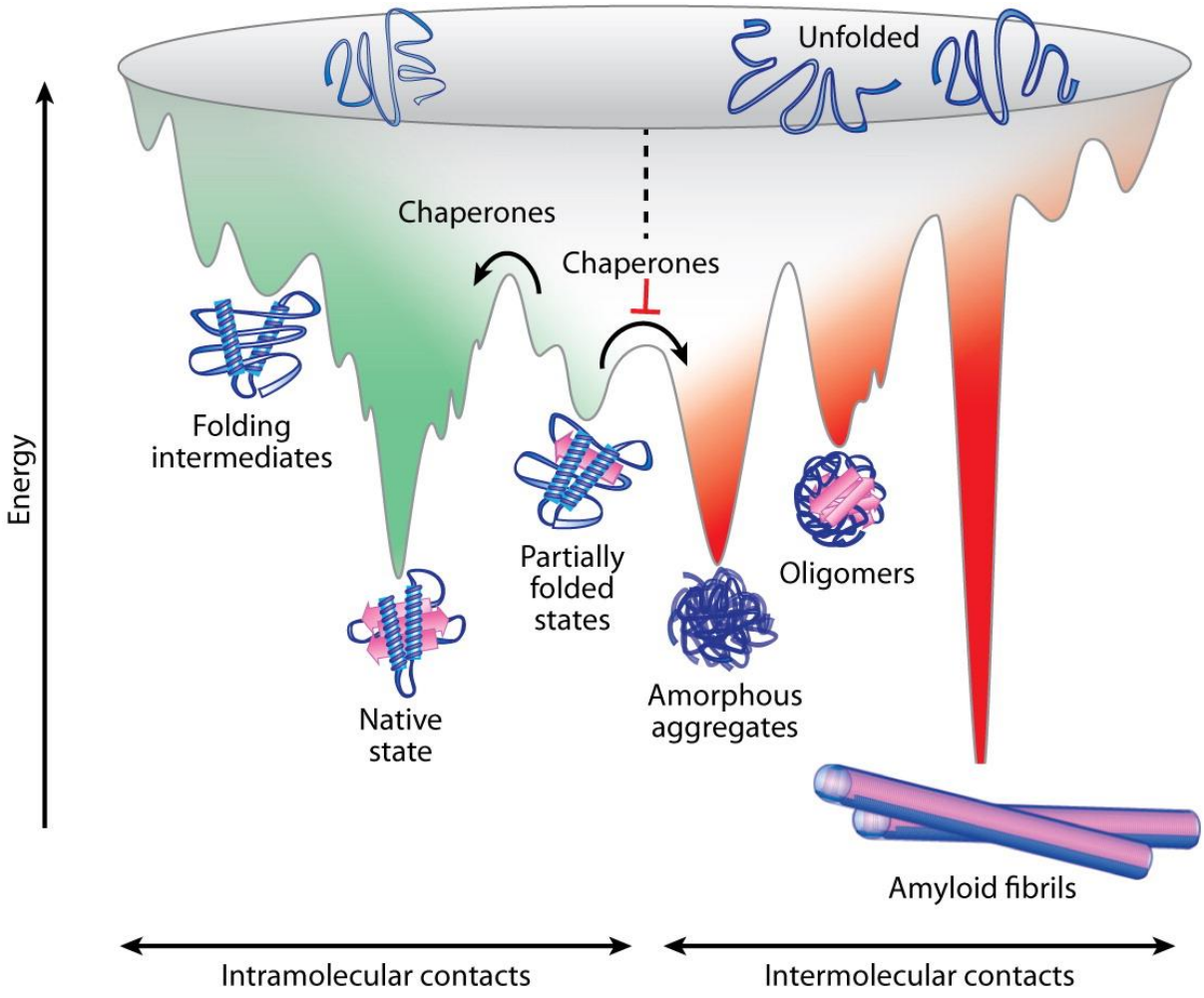

Figure 1-1. The protein folding energy landscape. Unfolded proteins have a higher free energy than folded proteins. Proteins sample multiple conformations as they fold towards the native state (green). Folding intermediates and partially folded states represent local energetic minima and are stabilised by energetically favourable but non-native contacts. The term "molten globule" is used to refer to protein states in which tertiary structure is disrupted but secondary structure is intact. Molecular chaperones mediate folding by assisting proteins in overcoming local energetic minima. Molecular chaperones also prevent non-native intermolecular contacts that lead to aggregated states (red). Image adapted from Kim et al. (2013).

1.2. Protein folding in the cell

In vitro protein folding experiments are commonly performed in dilute solutions using proteins purified to homogeneity (e.g. Anfinsen, 1961, Jackson and Fersht, 1991). Mechanistic information on protein folding can certainly be derived from these studies, but there are additional factors that greatly complicate folding *in vivo*.

Cells are crowded environments containing millions of molecules. The concentrations of protein, RNA, and DNA in *E. coli* have been estimated at 200 to 320 mg/mL, 75 to 120 mg/mL and 11 to 18 mg/mL respectively (Cayley et al., 1991 and Zimmerman et al., 1991). Excluded volume effects further increase the effective concentrations of molecules in cells, enhancing functional interactions between proteins but also creating an environment that can exacerbate protein folding errors (Ellis and Minton, 2006).

1.2.1. Misfolding, aggregation and disease

The native state of most proteins is only marginally stable, making them prone to misfolding and aggregation (Dobson et al., 1998). The folding free-energy landscape of proteins is rugged and contains local energetic minima that represent folding intermediate states; likely mixtures of native and non-native structure. Processes that slow the rate of protein folding, such as proline isomerisation, can lead to accumulation of these intermediate states. Protein folding intermediates require energy to escape these energetic minima and continue folding towards the native state. Hydrophobic residues, that would otherwise be buried, are presumably exposed to bulk solvent in these folding intermediates. This promotes association with other folding intermediates via hydrophobic interaction and can lead to the formation of amorphous aggregates, also called inclusions. Aggregation of proteins can also lead to the formation of ordered aggregates, called amyloid-like fibrils (Balchin et al., 2016). Accumulation of either inclusions or amyloid-like fibrils is cytotoxic and is linked to many human diseases including neurodegenerative diseases such as

Alzheimer's Disease, Parkinson's Disease, Huntington's Disease and prion diseases. Together, more than 40 human diseases are related to *in vivo* aggregation (Meric et al., 2017). To prevent this, cells have evolved multiple systems to mediate the correct folding of proteins, as well as an extensive protein quality control network to maintain proteostasis.

1.2.2. Co-translational protein folding

In protein synthesis, translation occurs over a longer time scale than folding. Hence proteins begin to fold during their synthesis on the ribosome (Thommen et al., 2017). An emerging polypeptide exits the ribosome through a channel that is ~ 100 Å in length and ~ 10 – 20 Å in diameter (Nilsson et al., 2015). The ribosomal exit tunnel restricts the conformational space that the nascent chain can sample (Kaiser et al., 2011). However, formation of both secondary structure elements and small domains has been observed to occur within this confined space (Bhushan et al., 2010 and Nilsson et al., 2015). The controlled vectorial nature (from N- to C-terminus) of translation has also been shown to enhance the folding of larger proteins, perhaps by allowing individual domains to fold separately in relative isolation (Frydman et al., 1999).

The nascent polypeptide interacts with a host of ribosome-associated chaperones such as trigger factor (TF) and Hsp70 in bacteria, or the nascent-polypeptide-associated complex (NAC) in eukarya (Craig et al., 2003 and Ott et al., 2015). While many proteins require no further assistance to reach their native states, proteins prone to misfolding typically require further processing by the protein quality control network, including the various chaperone families.

1.2.3. Cytosolic chaperone networks

In the 1980s, evidence emerged showing that many proteins required the presence of heat shock proteins (Hsp) for efficient folding even under non-stress conditions. These proteins became known as molecular chaperones (Ellis, 1996). Molecular chaperones are defined as proteins that mediate protein folding but are not present in the final structures of their substrates (Hartl, 1996).

Over several decades, a wealth of research has shown the essential roles of highly conserved families of chaperones in mediating different aspects of protein quality control (**Fig. 1-2** and Kim et al., 2013). Like transcription and translation, protein folding is prone to errors and as discussed, these errors can lead to disease. Molecular chaperones employ a variety of mechanisms to correct protein folding mistakes at different stages and prevent disease (Saibil, 2013).

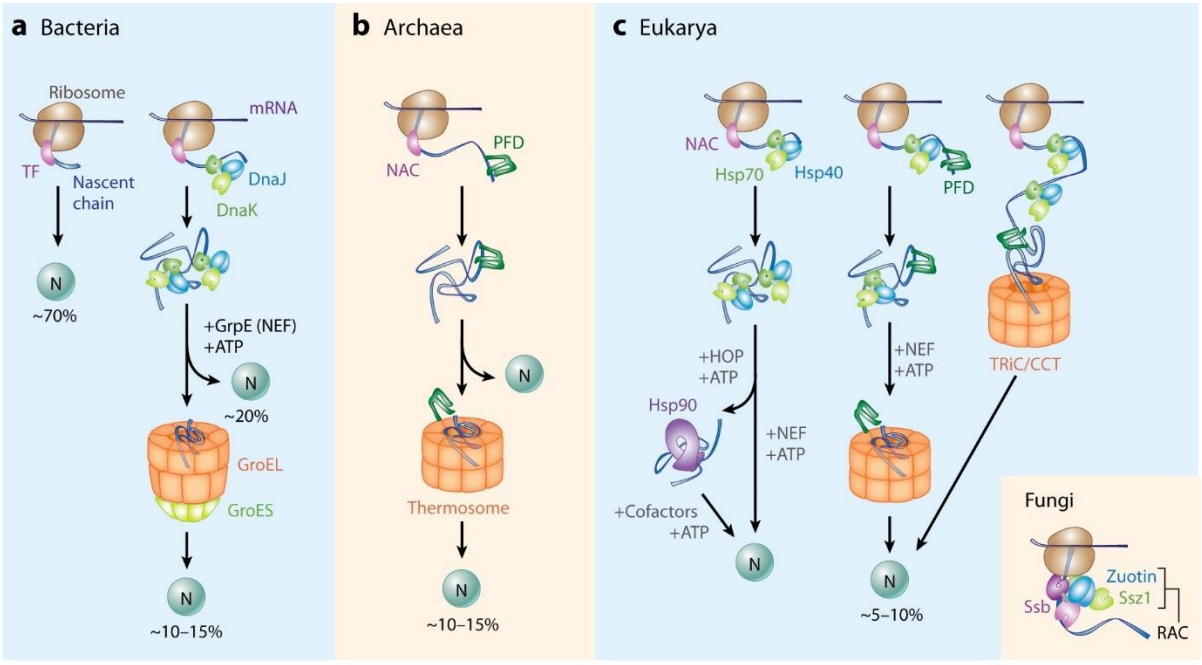

Figure 1-2. Cytosolic chaperone network of bacteria (a), archaea (b) and eukarya (c). Ribosome-associated chaperones (trigger factor [TF] in bacteria, nascent-chain-associated complex [NAC] in archaea and eukarya) mediate co-translational folding by binding to exposed hydrophobic regions on the emerging nascent chain. The 70 kDa heat shock proteins (Hsp70) chaperones (DnaK in bacteria, Hsp70 in archaea and eukarya) and their co-chaperones (DnaJ in bacteria, Hsp40 in archaea and eukarya) mediate protein folding by binding to the nascent chain or to non-native proteins. The 90 kDa heat shock proteins (Hsp90) chaperones and their co-chaperones mediate protein folding by binding non-native proteins. Partially folded or misfolded proteins may associate with the chaperonins for folding to the native (N) state (GroEL-GroES in bacteria, thermosome in archaea, tailless complex polypeptide-1 [TCP-1] ring complex

(TRiC)/chaperonin-containing TCP-1 [CCT] in eukarya). Percentages represent the approximate protein flux through the different chaperone networks. Image adapted from Kim et al. (2013).

The crucial role of chaperones in protein folding may appear at odds with Anfinsen's thermodynamic hypothesis. If the native three-dimensional structure of a protein only depends on its amino acid sequence, why do many proteins require chaperones to reach their native states *in vivo*? This can be reconciled by the differences between *in vitro* and *in vivo* protein folding. In a dilute solution, proteins fold to their native states in virtual isolation and off-target folding pathways are less likely. However, as discussed earlier, due to the (typically) warmer and more concentrated cellular environment, it quickly becomes kinetically favourable for proteins to follow off-target folding pathways that lead to misfolding or aggregation *in vivo*. Thus, molecular chaperones do not alter a substrate protein's native state; they only alter the folding pathway that it follows. Proteins more dependent on molecular chaperones are typically the same ones that are more prone to misfolding and aggregation.

1.2.3.1. The Hsp70 system

Pelham (1986) speculated that members of the Hsp70 family play a role in protein folding and assembly. Since then, Hsp70s have been shown to have a central role in the cytosolic chaperone network (Kim et al., 2013). Hsp70 proteins are ubiquitously expressed and are the most abundant cellular chaperones. Like all chaperones, Hsp70s recognise and bind to non-native proteins, however they also interact with a small set of native proteins to regulate their activity, stability, or assembly. Hsp70s have other functions, for example in the translocation of non-native proteins through organelle membranes (Mayer, 2013).

A host of co-chaperones interact with members of the Hsp70 family to mediate protein folding. Most of these co-chaperones are members of the Hsp40 family, called J-domain proteins in

bacteria. The primary role of co-chaperones is to recruit non-native substrates to Hsp70s and to enhance their intrinsically low ATPase activities (Saibil, 2013).

Hsp70s are monomeric and consist of two domains, a 43 kDa N-terminal nucleotide-binding domain (NBD) and a 27 kDa C-terminal substrate-binding domain (SBD), connected by a highly conserved flexible linker (**Fig. 1-3a** and Mayer, 2013). Investigation into the substrate specificity of Hsp70 showed that the SBD recognises a motif that is common to many proteins. The motif consists of a short stretch of 4 – 5 hydrophobic residues flanked by regions enriched in positively charged (basic) residues (Rudiger, 1997). In the native state of Hsp70 substrates, these motifs are typically buried in the hydrophobic core. In non-native states they become exposed, explaining the ability of Hsp70 to selectively bind non-native over native structures.

Hsp70 mediates folding through an ATP-driven cycle between open and closed states (**Fig. 1-3b**). In the first state, when ATP has bound, affinity for substrate protein is low; ATP hydrolysis converts Hsp70 to a closed state in which affinity for substrate protein is increased by 9 – 400-fold (Mayer, 2013). The intrinsic rate of ATP hydrolysis of Hsp70 is very low. (Mayer, 2013). Upon substrate binding to Hsp70, the ATPase rate is stimulated, and subsequent hydrolysis of ATP traps the substrate in the high-affinity, ADP-bound state of Hsp70.

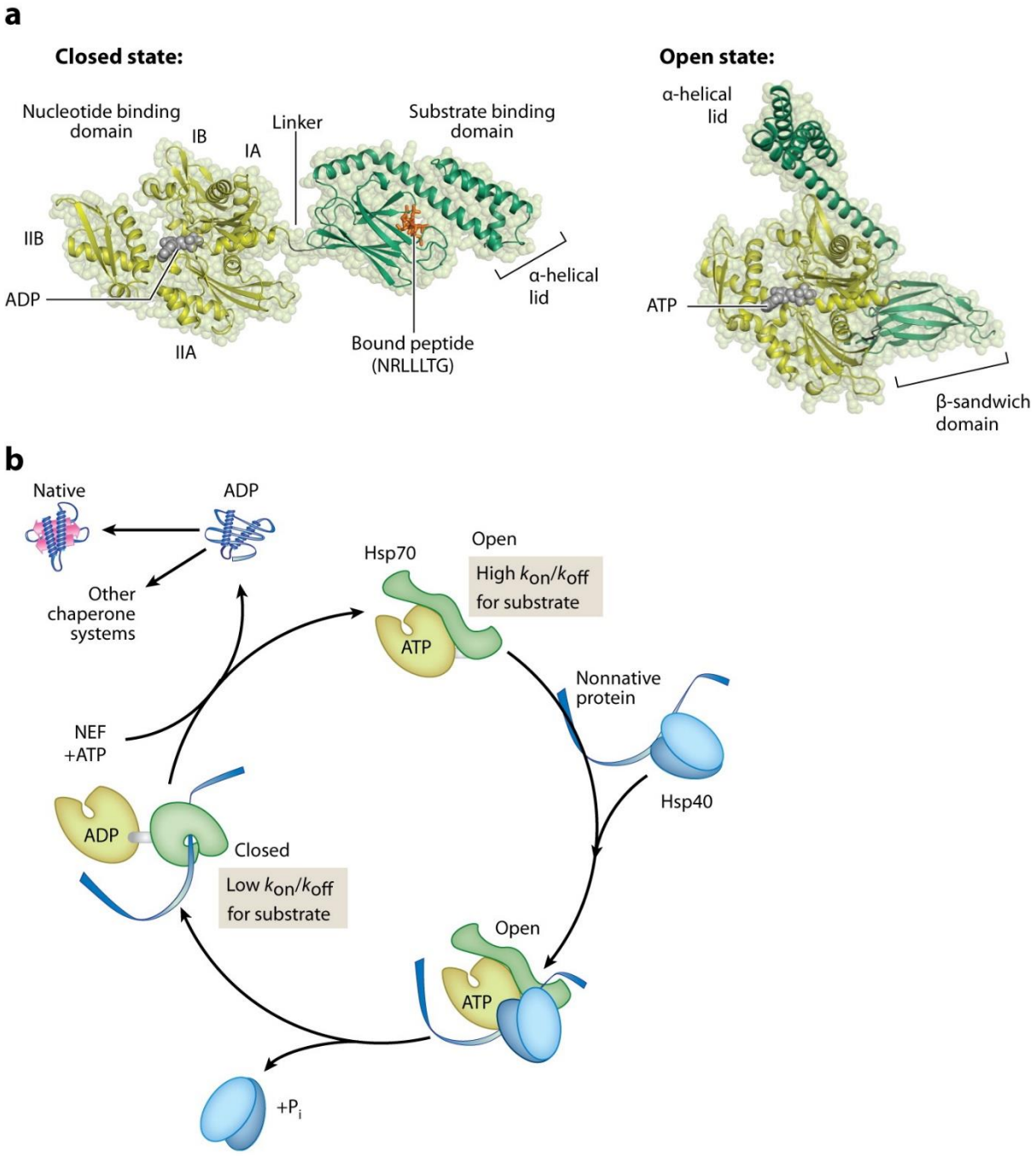

Figure 1-3. Structure (a) and reaction cycle (b) of Hsp70 (DnaK in bacteria). (a) Hsp70 consists of two domains, an N-terminal nucleotide-binding domain (NBD) and a C-terminal substrate-binding domain (SBD). The two domains are connected by a highly conserved linker. In the closed ADP-bound state (left structure, *E. coli* DnaK [PDB: 2KHO]), the NBD is separated from the SBD by the linker. The substrate peptide (NRLLLTG) is bound in the pocket of the β-sandwich domain. The α-helical lid is closed over the substrate binding pocket. In the open ATP-

bound state (right structure, yeast Sse1 [PDB: 2QXL]), subdomain IA of the NBD interacts with the β-sandwich domain of the SBD. Subdomains IA and IB of the NBD interact with the α-helical lid of the SBD. (**b**) ATP binding initiates the cycle by stabilising the open state of Hsp70. Non-native substrate is recruited to Hsp70 by Hsp40 co-chaperones. The stabilised open state has high on and off rates for non-native substrate. Both ATP-binding and Hsp40-binding stimulate ATP hydrolysis on Hsp70. ATP hydrolysis drives closing of the α-helical lid of the SBD. The ADP-bound closed state has low on and off rates for non-native substrate. Nucleotide-exchange factors (NEF) stimulate release of ADP and binding of ATP to Hsp70. ATP binding releases substrate from Hsp70. Image adapted from Kim et al. (2013).

1.2.3.2. The Hsp90 system

The Hsp90s are a class of chaperones expressed under non-stress conditions and upregulated under stress conditions. Hsp90s bind non-native proteins to prevent their aggregation, they also have an obligate role in the folding and maturation of certain proteins. Steroid hormone receptors and protein kinases are the best studied examples of Hsp90 substrates, the former of which requires Hsp90 for efficient loading of ligand (Saibil, 2013).

Members of the Hsp90 family form elongated dimers (**Fig. 1-4a**); each monomer contains three domains, a highly conserved N-terminal ATP-binding domain, a central substrate-binding and regulatory domain, and a C-terminal dimerization domain that also recruits numerous co-chaperones through its conserved Met-Glu-Glu-Val-Asp motif.

Like Hsp70, Hsp90 dimers undergo ATP-driven cycles between open and closed states (**Fig. 1-4b**). The open conformation of an Hsp90 dimer is V-shaped, each arm of the V comprising a single monomer. Non-native substrate binds between the two arms and causes the N-terminal domains to clamp together, trapping the substrate in the closed conformation of Hsp90. This extensive structural rearrangement is driven by ATP binding and hydrolysis, phosphorylation of

specific residues and interactions with multiple co-chaperones. The closed conformation of the dimer is compact, caused by the N-terminal domains twisting around each other. Subsequent ATP hydrolysis opens the dimer and releases the substrate protein in its active state.

Mechanistic details of substrate binding are poorly defined for Hsp90, with current models suggesting a lining of hydrophobic residues on the interior of the opened state, forming the substrate binding site. A recent cryoEM reconstruction of closed Hsp90 bound to its co-chaperone Cdc37 and the protein kinase Cdk4 revealed that Hsp90 and Cdc37 trap the kinase in a partially unfolded state in which its N-lobe is threaded through the chaperone. One of the apparent roles of Cdc37 is to interact with Hsp90 by mimicking the native interactions of the Cdk4 N-lobe. A suggested role for Cdc37 in the complex is that of quality-control; the co-chaperone may only dissociate from Hsp90 upon proper folding of the Cdk4 N-lobe (Verba et al., 2016).

Non-native proteins unable to use the Hsp70 or Hsp90 systems to reach their native state will instead use Hsp60 system, called GroEL in bacteria (Kim et al., 2013).

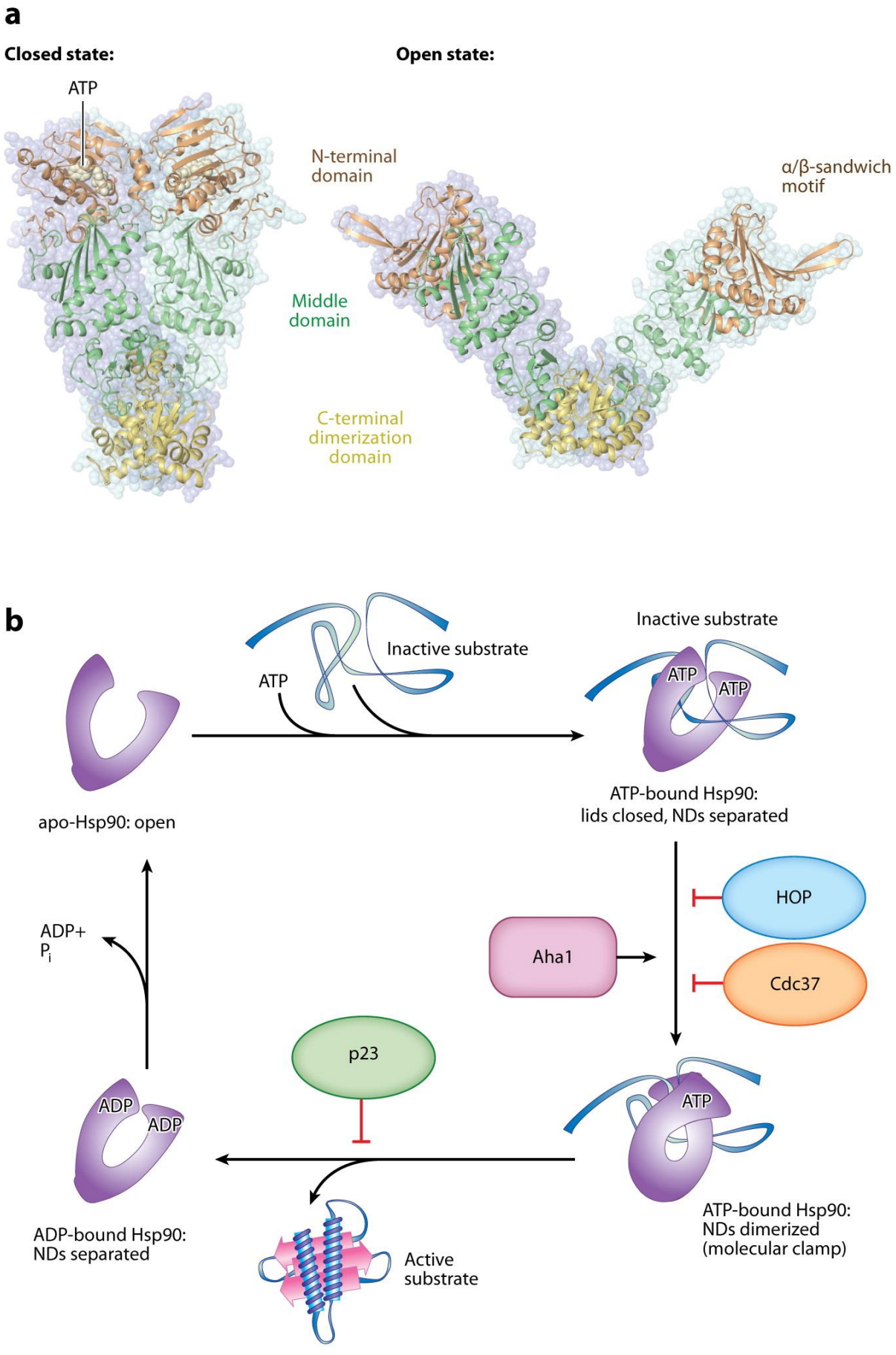
a
Closed state:
Open state:
ATP
N-terminal domain
α/β-sandwich motif
Middle domain
C-terminal dimerization domain
b
Inactive substrate
ATP
Inactive substrate
ATP
ATP
apo-Hsp90: open
ATP-bound Hsp90: lids closed, NDs separated
HOP
Aha1
Cdc37
ADP+ P_i
p23
ATP
ADP
ADP
ADP-bound Hsp90: NDs separated
Active substrate
ATP-bound Hsp90: NDs dimerized (molecular clamp)

Figure 1-4. Structure (a) and reaction cycle (b) of Hsp90. (**a**) Hsp90 is a homodimer, each Hsp90 monomer consists of three domains, an N-terminal domain (NTD) that binds ATP, a substrate-binding and regulatory middle domain (MD) and a C-terminal dimerisation domain (CTD). The C-terminal domain also recruits co-chaperones through its conserved Met-Glu-Glu-Val-Asp motif. The NTD and MD are connected by a charged, flexible linker. In the closed state (left structure, yeast Hsp90 [PDB: 2CG9]), the NTD of each monomer are twisted around each other, acting as a molecular clamp on the substrate. In the open state (right structure, *E. coli* Hsp90 [PDB: 2IOQ]), the dimer adopts a V-shaped conformation. (**b**) Inactive substrate binds to Hsp90-ATP. ATP binding to Hsp90 induces large conformational rearrangements that lead to NTD dimerisation and clamping of bound non-native substrate. In the closed state ATP hydrolysis is stimulated. ATP hydrolysis leads to dissociation of the NTDs and release of active substrate, followed by ADP and inorganic phosphate. Co-chaperones such as Hsp90 organising protein (HOP) and Cdc37 inhibit ATP hydrolysis and act as a quality control mechanism. Activator of Hsp90 ATPase (Aha1) enhances the rate of ATP hydrolysis. The cofactor p23 stabilises substrate-bound Hsp90 to slow the release of substrate from Hsp90. Image adapted from Kim et al. (2013).

1.3. The bacterial chaperonins: GroEL and GroES

GroEL and GroES belong to a class of molecular chaperones called the chaperonins, originally named by John Ellis and colleagues (Hemmingsen et al., 1988). The name GroE originates from studies of the bacteriophage λ. Mutations in a specific bacterial gene blocked λ growth (gro) and led to aggregation of phage heads. Compensatory mutations were mapped to the phage gene λE, thus the bacterial gene was named groE (Georgopoulos, 1992). Finally, L and S correspond to the large (57.3 kDa) and small (10.4 kDa) subunits of GroE.

1.3.1. Identification of GroEL as a molecular chaperone

After the discovery of the role of bacterial GroE in phage assembly (Georgopoulos et al., 1973), it was sixteen years before its role in protein folding was uncovered (Goloubinoff et al., 1989). Hsp60 had previously been identified as a heat shock response protein, but interestingly, was also found to be expressed under non-stress conditions (McMullin and Hallberg, 1987). Further study tentatively suggested that the primary function of Hsp60 under non-stress conditions was to mediate refolding of unfolded proteins following their import into mitochondria (McMullin and Hallberg, 1987). Subsequent bioinformatic analysis showed a high (> 50%) sequence identity between mitochondrial Hsp60, bacterial GroEL, and a chloroplast protein important for the assembly of Rubisco. This confirmed an evolutionary link between the three proteins and led to their classification as chaperonins (Hemmingsen et al., 1988).

1.3.2. Uncovering the cellular role of GroEL

Hartl and colleagues had shown that following import into mitochondria, unfolded dihydrofolate reductase (DHFR) did not refold spontaneously as had been observed *in vitro*. Instead, DHFR associated with mitochondrial Hsp60. Refolding and release of native DHFR then occurred in an ATP-dependent manner (Ostermann et al., 1989). *In vitro* refolding experiments by Lorimer and colleagues demonstrated a similar function for *E. coli* GroEL and its cochaperonin, GroES (Goloubinoff et al., 1989). Purified *Rhodospirillum rubrum* Rubisco was unfolded in a high concentration of urea or guanidine-HCl then diluted out of denaturant into a GroEL-containing buffer. Binary complexes of GroEL and non-native Rubisco monomers were formed. Refolding experiments carried out at various protein concentrations suggested two competing pathways for unfolded Rubisco upon dilution from its denaturant; i) irreversible aggregation or ii) formation of a GroEL-Rubisco binary complex.

Recovery of Rubisco activity was not possible in the absence of GroEL and could only be achieved by the addition of both GroES and ATP to preformed GroEL-Rubisco binary complexes. The authors suggested that because GroEL only recognised the non-native state of Rubisco, the specificity of the interaction must arise from some structural property that is present in the non-native state and absent in the native state. They also argued that because *E. coli* GroEL did not evolve to interact with *R. rubrum* Rubisco, this property is not likely to be specific to Rubisco alone.

1.3.3. Chaperonins in other domains of life

Since these early discoveries a wealth of research has been published concerning the chaperonins. They were found to be essential for survival under all conditions (Fayet et al., 1989). In the early 1990s a second distinct family of similarly essential chaperonins was discovered in archaea and in the eukaryotic cytosol, indicating that chaperonins were essential across all domains of life (Trent et al., 1991). Chaperonins have since been divided into two groups. Group I chaperonins are found in bacteria (GroEL), mitochondria (Hsp60) and chloroplasts (Chloroplast chaperonin 60 [cpn60]/Rubisco-subunit binding protein). Group II chaperonins are found in the eukaryotic cytosol (tailless complex polypeptide-1 [TCP-1] ring complex (TRiC) or chaperonin-containing TCP-1 [CCT]) and archaea (thermosome) (Lopez et al, 2015).

1.3.4. Early structural studies of GroEL-GroES

Early negative stain electron microscopy revealed the broad architecture of GroEL. Both the bacterial (Hohn et al., 1979 and Hendrix, 1979) and chloroplast (Pushkin et al., 1982) proteins formed the same double-ringed complex, with each ring consisting of seven radially arranged subunits. Similarly, when Hsp60 was examined by negative-stain EM, the same cylindrical structure was seen (McMullin and Hallberg, 1988). Further negative stain EM coupled with image processing indicated a back-to-back arrangement of the two GroEL rings (Zwickl et al., 1990). The first images of GroEL-GroES showed an asymmetric complex in which one GroES heptamer bound

to one end of a GroEL tetradecamer (Saibil et al., 1991). Later work began to reveal in more detail some of the conformational changes that GroEL underwent (Langer et al., 1992). Upon GroES binding to one end of GroEL a conformational change was triggered in the opposite end (*trans* ring), preventing a second GroES from binding there. This suggested that GroEL could only associate with one GroES oligomer at a time and that its cycle was asymmetric.

The first full three-dimensional view of a chaperonin came from a negative-stain EM reconstruction of *Rhodobacter sphaeroides* GroEL (Saibil et al., 1993). Although the resolution was limited, the two main domains of GroEL subunits were visible, as well as the back-to-back arrangement of the GroEL rings. In addition, some conformational changes upon ATP binding were evident. Subsequent work using cryoEM examined these ATP-driven changes in progressively greater detail (Chen et al., 1994, Roseman et al., 1996 and Clare et al., 2012).

1.3.5. Architecture of GroEL

The first crystal structure and atomic model of GroEL revealed a cylindrical complex 145 Å in height and 135 Å in diameter (Braig et al., 1994). Two seven-membered back-to-back rings surround a central cavity ~45 Å in diameter (**Fig. 1-5a** and **Fig. 1-5b**). The central cavity of GroEL is divided into two halves at the equatorial plane by the disordered 23-residue C-terminal tails of the fourteen GroEL subunits. The C-terminal tails are highly conserved and are important for correct protein folding inside the chaperonin cavity (Tang et al., 2006 and Tang et al., 2008). Each GroEL subunit is divided into three domains, i) an apical domain at the ends of the complex, ii) an equatorial domain at the waist of the cylinder, and iii) an intermediate domain linking the other two with hinge-like connecters (**Fig. 5c** and Braig et al., 1994). Each domain fulfils a distinct role. Helices H and I (**Fig. 1-5c**) of the apical domains form a hydrophobic collar around the inner surface of the cavity and act as the primary binding sites for both non-native polypeptide and GroES (**Fig. 1-5d** and **Fig. 1-5e**). The equatorial domains form the intra- and inter-ring contacts between GroEL subunits and hold the ATP binding sites (**Fig. 1-5c** and **Fig. 1-5f**). The connecting

intermediate domains serve to transmit conformational changes from the nucleotide-binding equatorial domains to the substrate-binding apical domains during the GroEL reaction cycle.

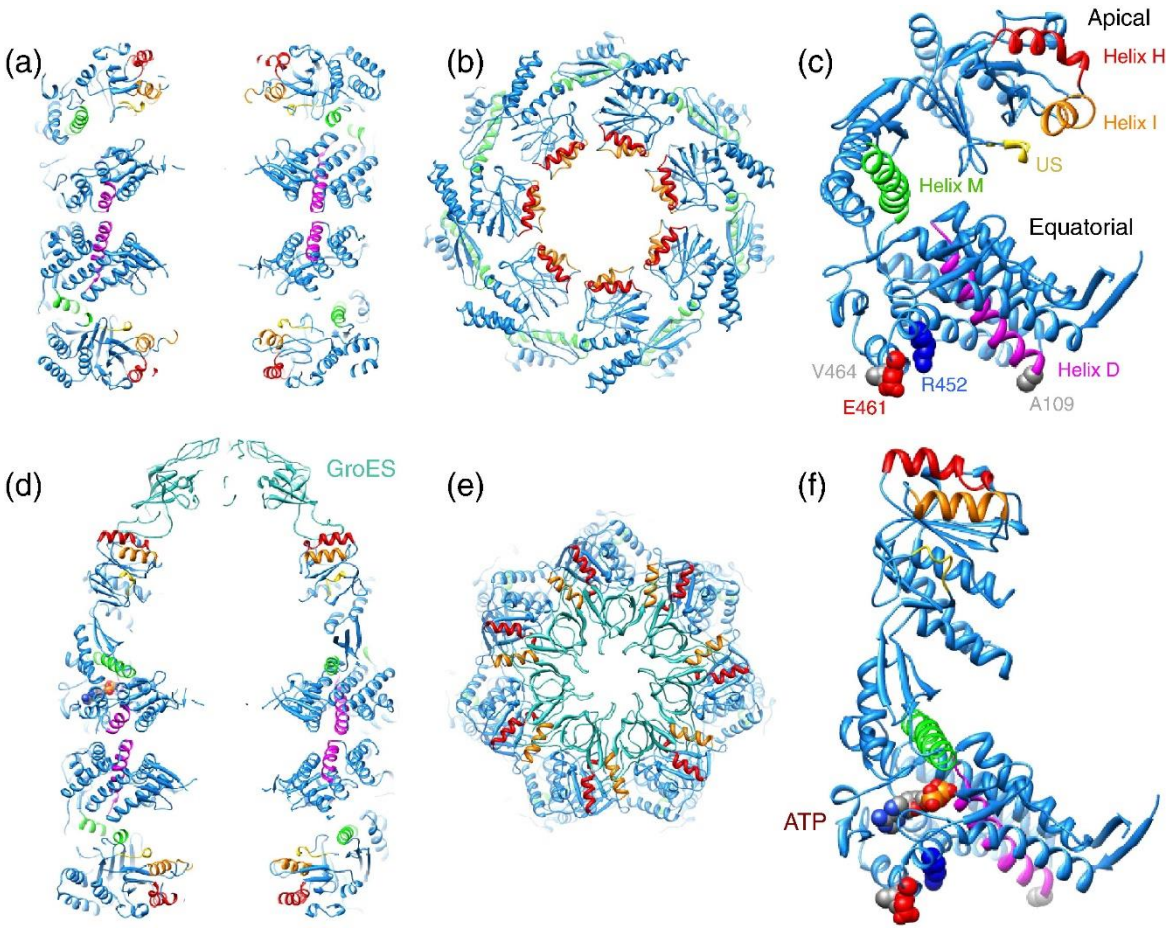

Figure 1-5. Architecture of GroEL and GroEL-GroES. Structures of GroEL and GroEL-GroES complexes. (**a**) Sliced-through side view of GroEL (PDB: 1OEL). (**b**) End view of GroEL. (**c**) GroEL subunit. Apical and equatorial domains are labelled. (**d**) Sliced-through side view of GroEL-GroES (PDB: 1SVT). (**e**) End view of GroEL-GroES. (**f**) GroEL subunit of GroEL-GroES and bound ATP. Highlighted areas are helix H (red), helix I (orange), underlying segment (yellow), helix M (green), helix D (purple), residues forming inter-ring contacts (V464, E461, R452 and A109). Image adapted from Saibil et al. (2013).

1.3.6. Architecture of GroES

GroES assembles into a heptameric dome-shaped ring 30 Å in height and 75 Å in diameter (Hunt et al., 1996). Subunits of GroES have mobile loop extensions that are disordered in solution but stabilise upon hydrophobic interaction with helices H and I of GroEL subunits (**Fig. 5d**, Fenton et al., 1994). The seven GroES subunits surround a small central cavity that becomes continuous with the GroEL cavity upon formation of a GroEL-GroES complex (**Fig. 1-5d** and **Fig. 1-5e** and Xu et al., 1997). GroEL undergoes conformational changes upon GroES binding (**Fig. 1-5d-f**). GroES can be thought of as a proteinaceous lid for the folding chamber. Its role may not be merely passive; however, a directed-evolution study revealed that mutations of a cavity-facing Tyrosine residue in GroES increased the folding rate of GFP but attenuated folding of some natural substrates of GroEL (Wang et al., 2002).

1.3.7. The GroEL-GroES ATP-driven reaction cycle

1.3.7.1. In the absence of substrate

GroEL undergoes cycles of ATP binding, hydrolysis, and release even in the absence of substrate protein. ATP-binding to GroEL exhibits positive cooperativity between subunits of a ring (intra-ring) and negative cooperativity between rings (inter-ring). Data from mass spectrometry supports the idea of an equilibrium between a low-affinity (T) and a high-affinity (R) state of GroEL for ATP (Dyachenko et al., 2013). Binding of ATP to one of the GroEL rings shifts the equilibrium of that ring towards the R-state and induces extension and rotation of the apical domains, promoting binding of GroES to form a *cis* ring (Clare et al., 2012). Hydrolysis of ATP to ADP in the *cis* ring increases affinity for ATP in the *trans* ring. Subsequent binding of ATP to the *trans* ring then releases ADP and GroES from the *cis* ring, restarting the cycle from the newly formed ATP-bound *cis* ring.

1.3.7.2. In the presence of substrate

A critical feature of the chaperonin mechanism is the ability to recognise and bind non-native substrate, but not substrates in their native state. This interaction is mediated by a hydrophobic groove located between two cavity-facing helices (H and I) on each apical domain of GroEL (Fenton et al., 1994). An underlying loop segment is equally important in mediating binding and folding of non-native substrate (Fenton et al., 1994). Exposed hydrophobic stretches of non-native substrate bind multivalently to these grooves (Farr et al., 2000). CryoEM reconstructions of GroEL-substrate complexes appear to show bound substrate adopting an intermediate folding state (Weaver et al., 2017) or perhaps sampling several different intermediate states (Elad et al., 2007).

Non-native substrate undergoes steps of conformational rearrangement as it is transferred from the apical domain of GroEL to the *cis* cavity of GroEL-GroES (**Fig. 1-6**). This has been shown by FRET experiments that monitored the average distance between the N- and C-termini of non-native Rubisco during binding and encapsulation (Lin and Rye, 2004). In the first step, non-native Rubisco bound to the *trans* ring of GroEL-GroES in a compact, kinetically trapped intermediate state (Lin et al., 2008). In the second step, ATP-induced conformational changes in the apical domain caused rapid expansion of Rubisco (Lin et al., 2008) consistent with a 'forced unfolding' mechanism (Shtilerman et al., 1999 and Stan et al., 2007). Forced unfolding may release substrate from a kinetically trapped state, pulling it back up the free-energy landscape. Binding of GroES releases non-native substrate into the *cis* ring cavity where it undergoes rapid conformational expansion followed by slow compaction, i.e. folding (**Fig. 1-6** and Lin et al., 2008).

Encapsulated substrate has a limited time (dependent on temperature) inside the GroEL-GroES cavity before all seven ATP are hydrolysed to ADP (**Fig. 1-6**). Following ATP hydrolysis in the *cis* ring, ATP binding to the opposite *trans* ring causes dissociation of GroES, ADP and substrate from the *cis* ring (**Fig. 1-6**). If the substrate has not folded to its native state during its residence

time inside the GroEL-GroES cavity, it is in close enough proximity to GroEL that it may rebind to a newly formed *cis* ring for repeated cycles of encapsulation and folding (Rye et al., 1999).

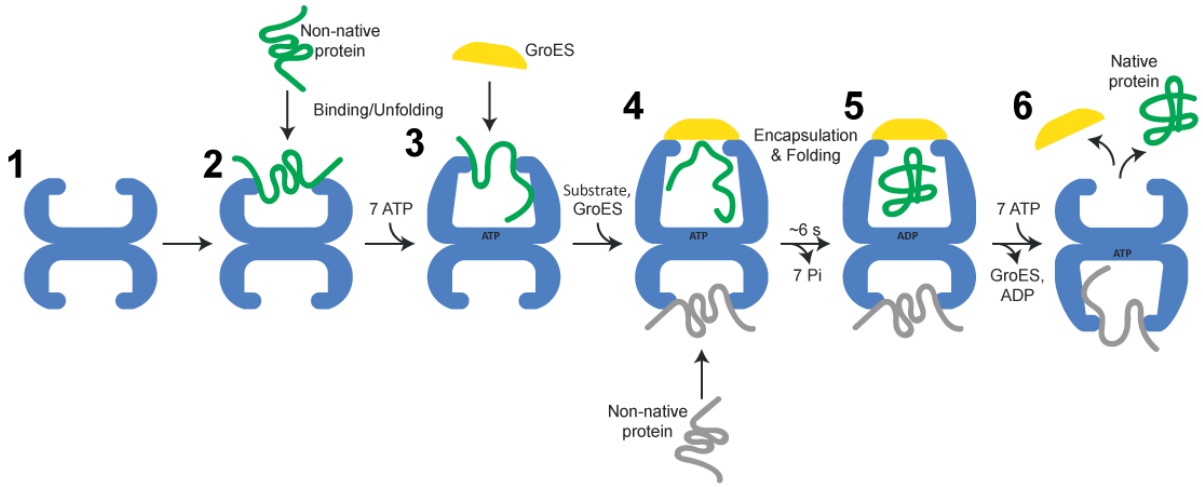

Figure 1-6. The GroEL-GroES reaction cycle. (**1**) Apo GroEL. (**2**) Non-native substrate protein binds to apo GroEL and undergoes conformational expansion. (**3**) ATP binding induces conformational changes in GroEL that might partially unfold bound substrate. GroES is recruited. (**4**) GroES binds to GroEL-ATP and triggers release of non-native substrate into the *cis* cavity. Additionally, a second non-native substrate binds to the *trans* ring of GroEL-GroES. (**5**) After ~6 s the seven molecules of ATP are hydrolysed to ADP. During this time, the protein folds in the *cis* cavity. (**6**) Following ATP hydrolysis in the *cis* ring, ATP binding to the *trans* ring induces dissociation of GroES, ADP and native substrate from *cis* ring. Cycles of binding, encapsulation and release then proceeds in cycles from step 3 to step 6. Note that complexes in step 3 and step 6 are functionally equivalent rotated 180°.

1.3.8. Structural basis of the GroEL-GroES reaction cycle

Substrate binding and the rigid body movements of GroEL during its reaction cycle have been characterised extensively by both cryoEM and X-ray crystallography.

1.3.8.1. Substrate binding

The reaction cycle of GroEL-GroES begins with the binding of non-native substrate to the apical domains of a GroEL *trans* ring. CryoEM reconstructions of GroEL-substrate binary complexes typically show a mass of density in one GroEL cavity (**Fig. 1-7**). However, to date no cryoEM reconstruction has resolved recognisable features (such as domains) of a bound non-native substrate. It is likely that substrate protein binds to GroEL in an ensemble of conformations which become averaged out during single-particle analysis.

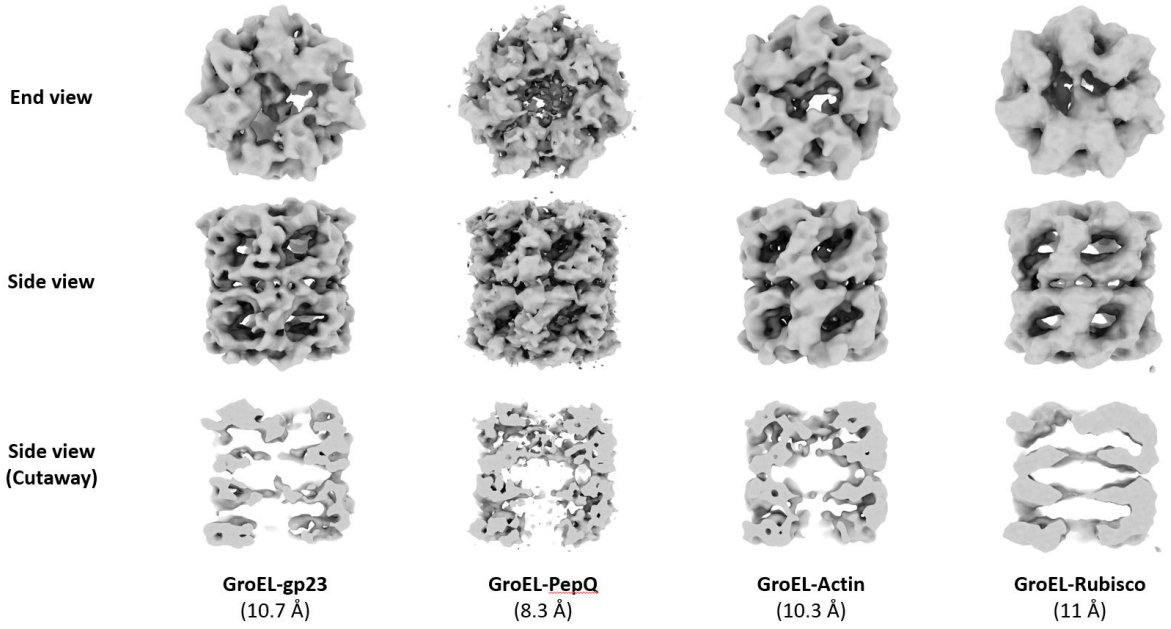

Figure 1-7. Published cryoEM reconstructions of GroEL-Substrate binary complexes. Maps downloaded from EMDB. GroEL-gp23 (EMD-1544), GroEL-PepQ (EMD-8316), GroEL-actin (EMD-0015) and GroEL-Rubisco (EMD-6726).

1.3.8.2. Cooperativity of ATP-binding

Irrespective of the presence of substrate polypeptide, ATP rapidly binds to apo GroEL or to the *trans* ring of GroEL-ADP-GroES in a cooperative fashion. CryoEM reconstructions of GroEL and GroEL-ATP (D398A) show that ATP binding induces a sequence of rigid body movements

with populated intermediate states termed Rs_1, Rs_2 and Rs-open (**Fig. 1-8** and Ranson et al., 2001 and Clare et al., 2012).

In Rs_1 (**Fig. 1-8b**), there is a 35° sideways rigid-body rotation of intermediate and apical domains about the lower hinge of the intermediate domain. Due to this movement, the catalytic residue Asp398 is brought into the ATP binding pocket, creating a network of hydrogen bonds. Associated with formation of the Rs_1 state is a reorganisation of salt bridges between neighbouring GroEL subunits. The downward movement of the intermediate domain breaks the Glu386 (intermediate) to Arg197 (apical) salt bridge and leads to formation of the Glu386 (intermediate) to Lys80 (equatorial) salt bridge. This is thought to communicate the ATP-binding status to neighbouring subunits (Ranson et al., 2001). Concurrent with this is the relocation of an apical-apical inter-subunit salt bridge from Lys207-Glu255 to Lys245-Glu255. This repositioning of inter-subunit salt bridges provides a model for the positive cooperativity of ATP-binding observed within GroEL rings (Ranson et al., 2001).

1.3.8.3. Forced unfolding

Formation of Rs_2 is marked by a slight rigid-body elevation of the apical domains. The state prior to GroES binding is Rs-open (**Fig. 1-8c**). In this state, apical domains move radially outward and elevate by a further 20°, breaking the inter-subunit apical salt bridges and physically separating apical domains from each other. The expansion of apical domains in this manner may form the structural basis for forced unfolding of the multivalently bound non-native substrate (Clare et al., 2012).

1.3.8.4. GroES-binding and substrate release

ATP-binding and the resulting conformational rearrangements prime GroEL for binding of GroES. In the Rs-open state, the hydrophobic substrate-binding helices face upward and in line with the mobile loops of GroES but remain accessible to the central cavity. Association between

the GroES mobile loops and the GroEL apical domain helices induces the largest rigid-body movements in the reaction cycle. Upon association, GroEL apical domains elevate to their final height and undergo a 100° clockwise twist to form the R-ES state (**Fig. 1-8d**). This drastic twist may constitute a power stroke that releases non-native substrate. If GroES binding and substrate ejection into the cavity happen simultaneously, this could explain why bound substrate does not escape prior to encapsulation (Clare et al., 2012).

During the cycle, both the apical domain helices and the C-terminal tails interact with non-native substrate. Although the C-terminal tails are predominantly hydrophobic, a hydrophilic region also appears to be important for higher folding rates (Machida et al., 2008). CryoEM reconstructions have shown encapsulated substrate interacting directly with the C-terminal tails of GroEL (Chen et al., 2013 and Weaver et al., 2017).

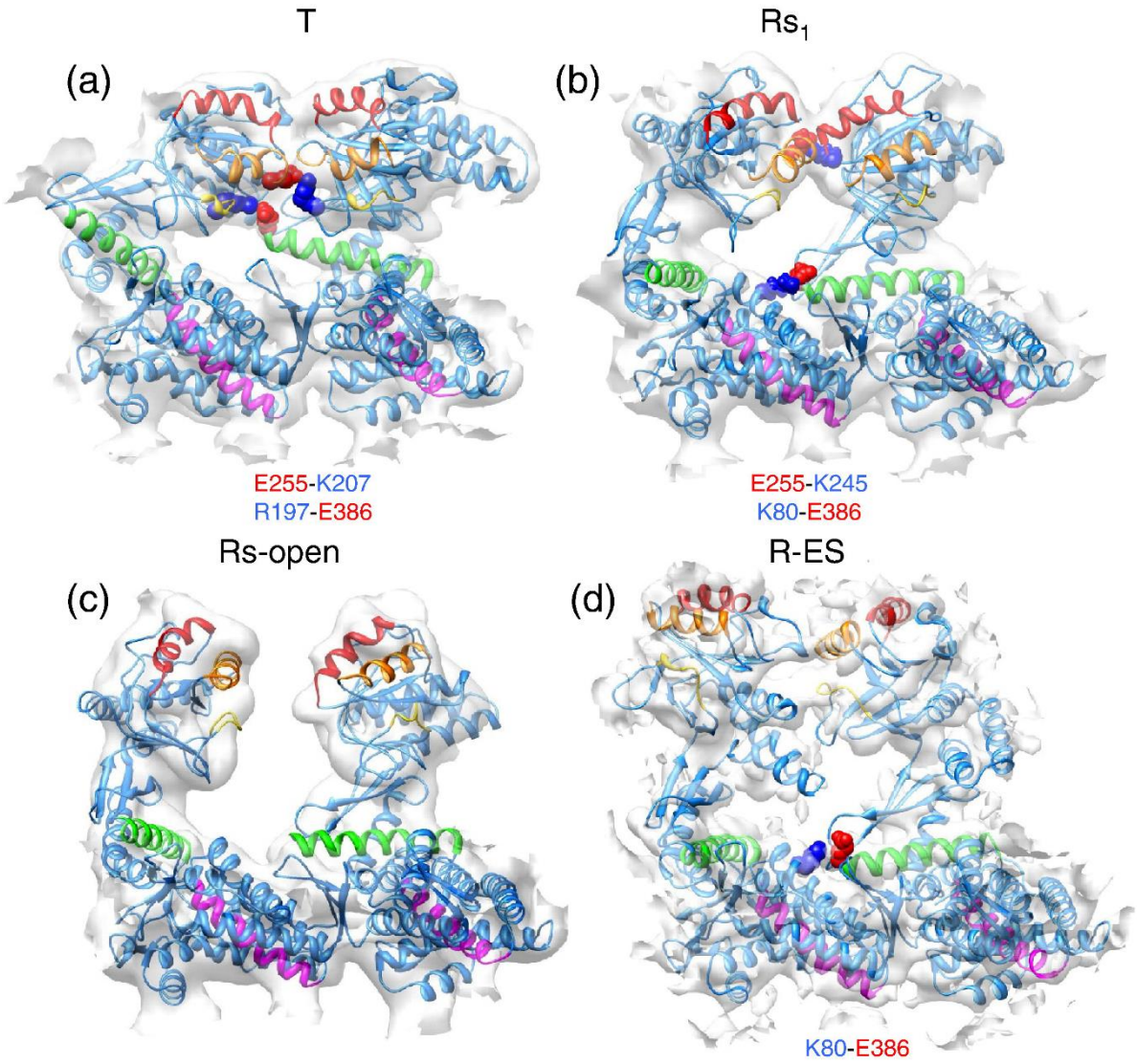

Figure 1-8. GroEL conformational states. CryoEM maps and fitted atomic structures of two adjacent subunits of (**a**) apo-GroEL (T state), (**b**) GroEL-ATP_7 (Rs_1 state), (**c**) GroEL-ATP_7 (Rs-open state) and (**d**) GroEL in the GroEL-GroES complex (R-ES state). The subunits are viewed from inside the GroEL ring. Highlighted regions are coloured as in **Fig. 1-5**. Image adapted from Saibil et al. (2013).

1.3.9. Substrates of GroEL

Although GroEL interacts with a wide range of non-native substrates *in vitro*, it assists in the folding of only around 250 (~10%) *E. coli* proteins *in vivo*. Studies have assessed the GroEL-dependency of cytosolic proteins in *E. coli* and found that many substrates are only partially dependent on

GroEL for correct folding. An example is firefly luciferase, which is released from GroEL-GroES in a non-native conformation that acquires its native state only by association with the Hsp70 system (Weissman et al., 1996). Additionally, in a population of rhodanese molecules, a portion are released from GroEL as near-native intermediates that subsequently fold to the native state without further assistance from chaperones (Mayhew et al., 1996).

A proteome-wide analysis assigned *E. coli* proteins to three classes based upon their predicted GroEL-dependency for correct folding (Kerner et al., 2005). Proteins predicted to be GroEL-independent were assigned to class I, whereas proteins predicted to be partially dependent on GroEL were assigned to class II (Kerner et al., 2005 and Fujiwara et al., 2010). A subset of 84 proteins were predicted to be obligate GroEL substrates and were labelled class III by Kerner et al. (2005). A later study based on GroEL-depletion experiments found that only 49 of these 84 class III proteins were in fact GroEL-dependent *in vivo* (Fujiwara et al., 2010). Class III was therefore split into two classes: class III^+ (GroEL-dependent for solubility *in vivo*) and class III^- (GroEL-independent for solubility *in vivo*). The 49 members of class III^+ were then combined with 8 newly identified obligate substrates to form class IV. In both studies, fewer than half of all cytosolic *E. coli* proteins were assessed for GroEL-dependency, leaving the possibility that many chaperonin-dependent proteins have yet to be discovered (Fujiwara et al., 2010).

1.3.9.1. Molecular masses of GroEL substrates

Class IV GroEL substrates share certain structural and biochemical features (Fujiwara et al., 2010). The molecular mass of 95% of the identified class IV GroEL substrates is < 60 kDa with a sharp drop-off above ~50 kDa. This is consistent with the volume of the GroEL-GroES cavity, which limits the size of substrates that can be encapsulated (Sigler et al., 1998). However, the study also identified a small number of non-essential class IV GroEL substrates between 64 – 68 kDa (Fujiwara et al., 2010).

Substrates with a molecular mass above ~60 kDa are too large for encapsulation by a GroEL-GroES *cis* ring. However, studies have shown GroEL can enhance folding of several proteins larger than 60 kDa, including yeast mitochondrial aconitase (82 kDa; Chaudhuri et al., 2001) and *E. coli* maltodextrin glucosidase (69 kDa; Paul et al., 2007). In these studies, the large non-native substrate bound to the *trans* ring of a GroEL-GroES complex and was subsequently released into free solution. In such cases GroEL may act to unfold substrates, releasing them from kinetically trapped folding intermediates and rescuing them from aggregation (Lin et al., 2008). This is supported by the observation that non-encapsulating (*trans*-only) GroEL-GroES complexes retain the ability to enhance the folding rate of two classical obligate GroEL substrates, Rubisco and MDH (Farr et al., 2003).

1.3.9.2. Structural features common to GroEL substrates

A second common feature among GroEL substrates is their tendency to aggregate. In GroEL-depleted cells all but one of the 57 class IV substrates were highly prone to aggregation during translation (Fujiwara et al., 2010). This tendency to aggregate may be an *in vivo* determinant for GroEL dependency and may be related to the tertiary structure of substrate proteins. In support of this view is the strong enrichment of the $(\beta\alpha)_8$ triose-phosphate isomerase (TIM) barrel fold within class IV GroEL substrates (Houry et al., 1999 and Kerner et al., 2005). Although present in only 7% of the *E. coli* lysate proteins, TIM barrels were present in ~45% of class IV substrates (Fujiwara et al., 2010). Along with the TIM barrel, several other aggregation-prone structural folds were also overrepresented in class IV substrates (Fujiwara et al., 2010).

During spontaneous folding, TIM barrels populate an ensemble of kinetically trapped intermediates, leading to long folding times (Georgescauld et al., 2014). These intermediates may share certain structural features that confer GroEL dependence. However, not all TIM barrel proteins are GroEL-dependent. Recent approaches using machine learning have been unable to

reconcile this, failing to distinguish defining features of GroEL-dependent and GroEL-independent TIM barrel-containing proteins (Azia et al., 2012).

1.3.9.3. Evolution of GroEL substrates

In evolution, mutations are necessary for proteins to gain new functions. However, most mutations have a destabilising effect on protein structure (Tokuriki et al., 2008). Loss of thermodynamic stability can lead to accumulation of kinetically trapped folding intermediates, resulting in aggregation (Tokuriki and Tawfik, 2009a). Such mutations would therefore not be favoured by natural selection (DePristo et al., 2005). Hsp90 chaperones have been proposed to act as buffers against the destabilising effects of mutations, thereby facilitating the evolution of new protein functions (Rutherford and Lindquist, 1998). This hypothesis has been extended to the chaperonins (Fares et al., 2002 and Williams and Fares, 2010). In support of this, overexpression of GroEL and GroES has been demonstrated to stabilise mutants of otherwise GroEL-independent proteins as well as GroEL-dependent proteins (Tokuriki and Tawfik, 2009b and Mueller-Cajar and Whitney, 2008). Furthermore, bioinformatic analyses have highlighted a correlation between evolutionary rates and GroEL-dependency, with obligate GroEL substrates evolving 35% faster than GroEL-independent substrates (Bogumil and Dagan, 2012).

Orthologs of obligate GroEL substrates have been discovered in a small group of bacteria that lack the GroE gene, suggesting that these proteins have escaped chaperonin dependence during evolution (Georgescauld et al., 2014 and Ishimoto et al., 2014).

1.3.10. Mechanisms of GroEL-assisted folding

The question surrounding the exact mechanism of chaperonin-assisted protein folding has a long history (Fenton and Horwich, 1997 and Hayer-Hartl et al., 2016). Several models have been proposed, each of them falling into one of two broad categories, active or passive involvement. Given that aggregation and protein folding are competing pathways (Dobson, 2003), how do

chaperonins promote protein folding? Do they function exclusively to prevent aggregation? Or do they promote folding using active mechanisms?

1.3.10.1. Passive model

In the passive model, also called the Anfinsen cage model, chaperonins are proposed to enhance the rate of protein folding by simply sequestering substrates and passively preventing folding intermediates from aggregating (Jewett and Shea, 2009). A key aspect of the passive model is that following encapsulation, chaperonins do not alter the folding kinetics of substrate proteins (**Fig. 1-9**, Horst et al., 2007 and Tyagi et al., 2011). However, the repeat observation that chaperonins accelerate the rate of folding of some proteins conflicts with a purely passive model (Brinker et al., 2001, Chakraborty et al., 2010, Dahiya and Chaudhuri, 2013).

In studies showing that chaperonins increase the rate of folding, the passive model argues that aggregation is the cause of the slow spontaneous folding of substrate proteins, and that chaperonins merely act as 'infinite-dilution cages' to restore the rate of folding (Apetri and Horwich, 2008). Recent studies argue strongly against this, as discussed later (Georgescauld et al., 2014 and Weaver et al., 2017).

1.3.10.2. Active cage model

The active cage model (**Fig. 1-9**) proposes that the rate enhancement of folding is a direct result of the environment of the chaperonin cavity. Confinement of a substrate protein "smoothens" the free energy landscape by either blocking the formation of kinetically trapped intermediates, or facilitating their transition to the native state. One can imagine confinement in the GroEL-GroES cavity preventing formation of extended conformations of misfolded proteins and promoting more compact states. Supporting this, three specific properties of the GroEL-GroES cavity have been shown to be critical to accelerate folding of encapsulated substrates; i) the volume of the cavity (Tang et al., 2006), ii) the net-negatively charged cavity wall (Tang et al., 2006 and Gupta et

al., 2014) and iii) the cavity-protruding C-terminal tails (Tang et al., 2006 and Weaver and Rye, 2014).

Experiments by Georgescauld et al. (2014) on DapA, an *in vivo* obligate GroEL substrate, were designed to test the passive cage model by measuring rates of folding in the complete absence of aggregation. If folding was still enhanced by GroEL when no aggregation was present, it would be evidence for an active model. Refolding experiments of DapA were carried out at picomolar concentrations, at which prior single-molecule fluorescence experiments had confirmed the absence of any intermolecular association (i.e. aggregation or assembly of oligomers). The rate of GroEL-assisted refolding was shown to be > 50-fold higher than spontaneous refolding. In the absence of aggregation, this result cannot be reconciled with a purely passive cage model.

Additionally, hydrogen-deuterium exchange coupled to mass spectrometry (HDX-MS) was used to compare the spontaneous and chaperonin-assisted folding pathways of DapA. During spontaneous folding, the TIM-barrel of DapA formed in a concerted manner from an ensemble of dynamic folding intermediates, consistent with a high free-energy barrier and long search time (Georgescauld et al., 2014). In the presence of chaperonins, folding was less concerted, proceeding in a series of smaller steps from several local nucleation sites. This was consistent with a reduced entropic barrier and a smoothened free energy landscape. The same stepwise structure formation was also seen in a structural homologue of DapA that is GroEL-independent (Georgescauld et al., 2014). Together, these results argue in favour of an active model by demonstrating that chaperonins alter the folding kinetics of substrates independently of their role in aggregation-prevention.

Weaver et al. (2017) recently performed experiments designed to discriminate between the passive and active models. For PepQ, another obligate GroEL substrate, their results suggested that in the absence of aggregation, the rate of GroEL-assisted folding was higher than the rate of spontaneous folding, again favouring an active model over a passive model.

Additionally, the initial capture of PepQ resulted in unfolding that was at least partially dependent on an interaction between the substrate and the disordered C-terminal extensions of GroEL. In cryoEM reconstructions of PepQ-bound GroEL, substrate density was positioned deeper inside the cavity of wild-type GroEL than of a C-terminally truncated mutant GroEL. This supported an interaction between the GroEL C-terminal extensions and the PepQ folding intermediate (Weaver et al. 2017).

1.3.10.3. Iterative annealing model

The iterative annealing model (**Fig. 1-9**) proposes that ATP-driven cycles of substrate binding and release actively accelerate folding. In this model, obligate substrates are those most likely to form kinetically-trapped folding intermediates that cannot be easily rescued by thermal motion and thus require unfolding (Todd et al., 1996). In this model, folding to the native state may occur either inside or outside of the chaperonin cavity (Priya et al., 2013). This distinguishes it from both passive and active models, but conflicts with evidence showing that folding must occur inside the cavity in order to be accelerated (Brinker et al., 2001, Tang et al., 2006 and Tang et al., 2008).

Binding to the *trans* ring of GroEL-ADP-GroES or to apo-GroEL induces partial, passive unfolding of some substrates (Lin et al., 2008 and Weaver et al., 2017), possibly analogous to surface-mediated denaturation (Swain and Gierasch, 2005). Upon ATP binding to the *trans* ring of GroEL-GroES, an active, forced unfolding step has been observed to cause conformational expansion of bound substrate (Lin et al., 2008). Forced unfolding offers substrates an additional chance to fold and has been suggested as a main cause of the enhanced the rate of folding (Lin et al., 2013). This prediction was recently tested by Weaver et al. (2017). They showed that under conditions of increasing GroEL ATPase cycling rates, GroEL-GroES exhibited a higher rate of folding than SR1-GroES, a single-ring mutant that only undergoes one half-cycle. This suggested that repetitive cycling of GroEL, including unfolding of substrate, was required to increase the rate of folding of PepQ.

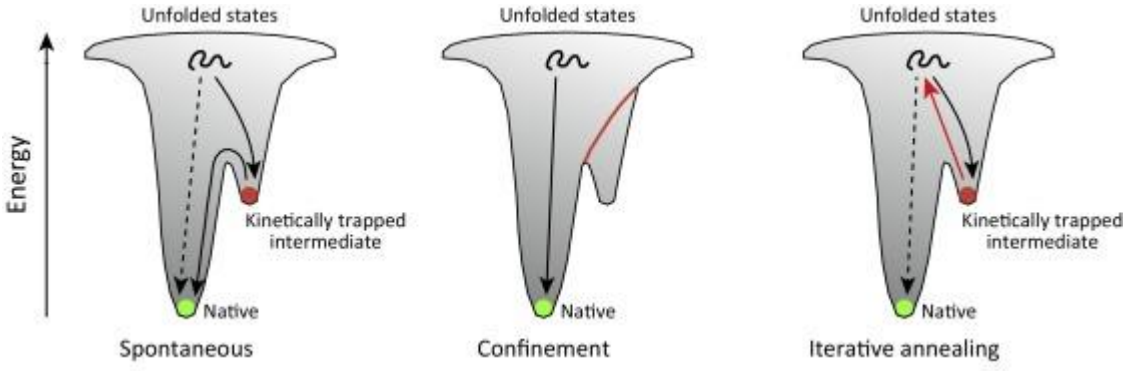

Figure 1-9. Models of chaperonin-assisted folding. Hypothetical folding energy landscapes for a protein that populates a kinetically trapped intermediate during spontaneous folding (left). The passive cage model suggests that chaperonins only prevent aggregation, allowing folding to proceed as it would spontaneously. The active cage model suggests that confinement in the GroEL-GroES cavity smoothens the folding energy landscape and population of kinetically trapped intermediates is avoided (middle). The iterative annealing model suggests that forced unfolding of kinetically trapped intermediates is responsible for accelerated folding (right). Image adapted from Hayer-Hartl et al. (2016).

1.3.10.4. A mixed model

Passive and active models are not entirely mutually exclusive. Neither are aspects of the iterative annealing and active cage models. The exact mechanism of chaperonin-assisted folding appears to be substrate-dependent, with GroEL acting as a catch-all solution. Upon encapsulation, most substrates certainly benefit from aggregation-prevention, but not all GroEL-substrates exhibit an increased folding rate. For example, rhodanese (~30 kDa) appears to use GroEL solely for prevention of aggregation (Brinker et al., 2001). It could be that rhodanese is too small to benefit from the effects of confinement (Tang et al., 2006), a conclusion that is consistent with mutational studies showing that a slight (2 – 4%) reduction in the volume of the GroEL-GroES cavity increases the folding rate of rhodanese by ~30% (Weaver and Rye, 2014). Furthermore, proteins that populate ensembles of kinetically-trapped intermediates may depend on forced unfolding (for

conformational rescue) followed by confined refolding in the chaperonin cavity (Georgescauld et al., 2014 and Weaver et al., 2017). A small subset of proteins with unusual topologies, such as knots, may also benefit from both passive and active mechanisms (Lim and Jackson, 2015).

1.4. Project aims

We aimed to understand the molecular mechanisms of assisted protein folding by the GroEL-GroES system. Previous structural studies of GroEL-substrate complexes were limited in resolution. Additionally, crystal structures of GroEL-substrate complexes do not resolve density for bound substrate (Shimamura et al., 2004 and Fei et al., 2014). By using the latest advances in cryoEM, we hope to characterise the interactions between GroEL and substrate at different steps of the GroEL reactive cycle. We opted to use Rubisco from *R. rubrum* as the GroEL substrate for all experiments. Rubisco is a large stringent substrate of GroEL and among the best studied.

Chapter II

2. CryoEM and single-particle analysis

This chapter introduces cryoEM and single-particle analysis, and provides an overview of the main experimental and computational principles behind the technique. It is based on lecture courses (EMBO 2017 Image processing for cryo-electron microscopy and LMB 2017 cryoEM course) and references (Orlova and Saibil, 2011, Scheres, 2012a, Scheres, 2012b, Cheng et al., 2015, Sigworth, 2015, Passmore and Russo, 2016 and Zivanov et al., 2018).

2.1. Techniques in structural biology

In structural biology there are several main techniques for determining the three-dimensional structures of proteins. These are X-ray crystallography, nuclear magnetic resonance (NMR) and cryogenic electron microscopy (cryoEM). X-ray crystallography has been the dominant technique in the field since the 1950s when John Kendrew and colleagues solved the first three-dimensional protein structure (Kendrew et al., 1958). Crystal structures now make up ~90% of the ~175,000 protein structures in the Protein Data Bank (EMBL-EBI, 2021) and RCSB.ORG, 2021). If a protein can be crystallised, its 3D structure can usually be solved. However, the ability to routinely produce diffraction-quality crystals is a significant bottleneck in X-ray crystallography. Crystallisation is often challenging as it requires a high concentration of purified protein and an exhaustive screening process to identify optimal crystallisation conditions. Proteins that are highly flexible or membrane-associated often fail to crystallise altogether (Deller et al., 2016 and Radivojac, 2004).

The main alternatives to X-ray crystallography do not rely on protein crystallisation. NMR can produce solution structures of proteins and provide information on protein dynamics making it a useful tool for studying flexible protein structures. NMR is mainly applicable to proteins smaller

than around 25 kDa. Above this size it becomes increasingly difficult and time-consuming to acquire high quality data (Frueh et al., 2013). However, with NMR it is becoming easier to study larger protein complexes (Schütz and Sprangers, 2020).

CryoEM encompasses several techniques that can provide structural information across a range of scales from the whole cell (using electron cryo-tomography) down to individual molecules (using single-particle analysis). Like X-ray crystallography, cryoEM has its own set of obstacles that hinder routine high-resolution structure determination.

2.2. Imaging biological specimens by transmission EM

The primary limit to high resolution structure determination by transmission EM is the radiation sensitivity of biological samples. Organic matter is composed of low atomic weight atoms, primarily hydrogen, carbon, nitrogen, and oxygen. The energy of covalent bonds in typical biological samples is around a few eVs (Baker and Rubinstein, 2010). However, the electron energies used in cryoEM experiments are 200 – 300 keV, several orders of magnitude higher. Thus, when biological samples are irradiated by an electron beam their chemical bonds begin to break. This generates free radicals that cause further secondary damage and ultimately destroy the sample (Glaeser, 1971). To limit radiation damage and preserve high resolution features, samples are imaged under "low-dose" conditions. Any sample inserted into an electron microscope must also be able to withstand the high vacuum of the column. Biological specimens would rapidly dehydrate if not fixed in some way.

2.3. Sample preparation for biological TEM

2.3.1. Biochemistry

Sample preparation should be optimised through biochemical methods prior to high resolution cryoEM. Protein purity can be evaluated by SDS-PAGE with Coomassie or silver-staining. Protein

complexes should be compositionally homogeneous as shown by native-PAGE, SEC-MALS or native mass spectrometry. Buffer conditions (e.g. pH, salt and detergent) should be optimised to maximise the stability of the complex while retaining its biological activity. Ideally complexes should be stable through a range of concentrations from 0.05 to 10 μM.

2.3.2. EM Grids

Samples for EM are applied to an EM grid in microlitre quantity (typically ~3 μL). An EM grid consists of a metal mesh that supports a thin film. Grid meshes for cryoEM are typically made from copper and have a mesh size of 200 – 400 lines per inch. The type of support film differs depending on application. Negative stain EM grids use a continuous layer of amorphous carbon. Grids for cryoEM have a perforated carbon film with regular hole size and spacing, and are usually referred to as holey grids. Other grid and film materials are available and more recently all-gold grids have been introduced (Russo and Passmore, 2014). In recent years, graphene derivatives such as graphene oxide have become a popular choice of grid support. Graphene oxide is more electron transparent than amorphous carbon. Holey grids can be made or purchased with different hole sizes and spacings. Common grids use 1.2 or 2.0 μm diameter holes and a 1.3 or 2.0 μm spacing between holes. Untreated EM grids are hydrophobic and are almost always subjected to glow discharge or plasma cleaning before use. This treatment decreases their hydrophobicity and allows the small volume of sample to be evenly distributed across the grid.

2.3.3. Negative stain EM

Negative staining (NS) is a quick and easy method of specimen preparation for EM (**Fig. 2-1A**). It is often used as a preliminary method to screen for sample properties such as homogeneity and particle distribution. In negative stain EM samples are applied to an EM grid with a continuous carbon film. Excess solution is removed and the specimen is embedded in a high molecular weight stain. The stain is commonly a heavy metal salt such as uranyl acetate (UA). Excess stain is removed

and the remaining thin layer dries and forms a cast around the specimen. The heavy metal stain increases specimen contrast and can withstand relatively high electron doses. Three-dimensional structures can be determined by negative stain EM, but their resolution is limited by the grain size of the stain used (e.g. ~20 Å for UA). Internal features are also not well-resolved in NS-EM since the stain does not penetrate inside the structure. Negatively stained samples can be prepared with minimal time and equipment meaning upwards of ~20 conditions can be screened in a single day.

2.3.4. CryoEM

Negative stain EM can introduce structural artefacts and is limited in resolution. To resolve the high-resolution features of biological samples it is necessary to image them in near-native conditions in the absence of stain. To preserve their native structures samples must remain hydrated and radiation damage must be minimised. CryoEM achieves this by i) embedding the sample in a thin layer of amorphous ice, ii) maintaining the sample at a low temperature during imaging and iii) limiting the total electron exposure.

2.3.5. Plunge-freezing

Plunge-freezing is the most widely used method of sample preparation for cryoEM. Sample is applied to an EM grid and excess solution is removed by blotting leaving the specimen suspended in a thin layer of aqueous buffer that is unsupported over holes. The grid is then rapidly plunged into a cryogen cooled to liquid nitrogen temperature (**Fig. 2-1B**). Commonly used cryogens such as liquid ethane have a very high thermal conductivity and cool the thin layer of buffer by ~10^4 °C/s (Dubochet et al., 1988). Rapid cooling freezes the water molecules in a glass-like amorphous state (vitreous ice) and avoids the formation of crystalline ice. Vitrified biological specimens remain hydrated preserving their native (or near-native) structures. Frozen grids must be stored and imaged at liquid nitrogen temperatures to ensure that the ice remains in a vitreous state. Devitrification occurs above temperatures of -137 °C and leads to formation of crystalline ice that

can damage the sample and interfere with imaging (Dubochet et al., 1988). Maintaining samples at liquid nitrogen temperature during imaging also offers a degree of cryoprotection and allows imaging to continue for longer (Baker and Rubinstein, 2010).

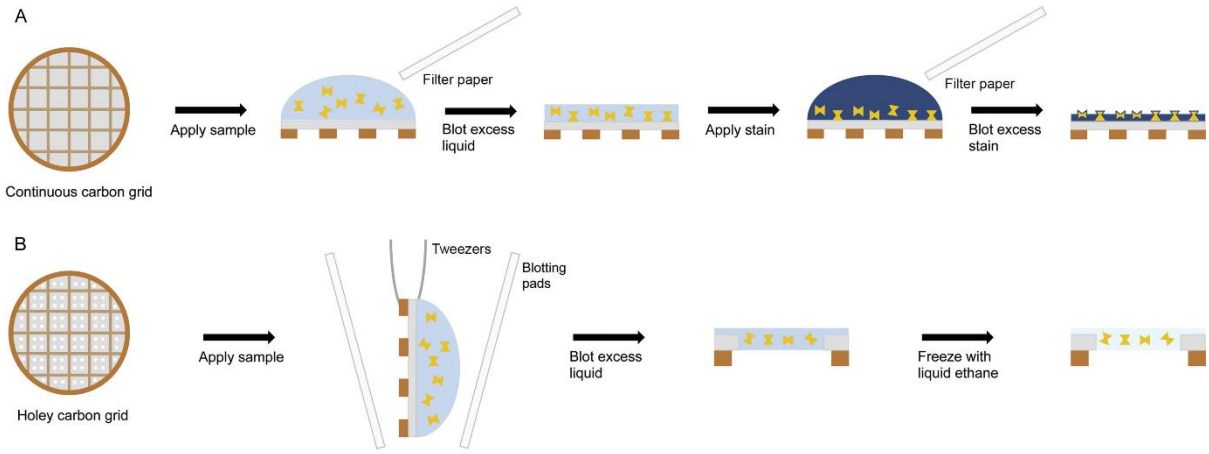

Figure 2-1. Grid preparation for negative stain EM and cryoEM. (A) In negative stain EM sample is applied to a grid with a continuous carbon layer and stained with a heavy metal salt. Excess liquid is removed by blotting and the stain dried. (B) In cryoEM sample is applied to a grid with a perforated film. Excess solution is removed by blotting and the grid is plunge-frozen. Biological specimens become suspended in a layer of vitreous ice.

2.3.6. Optimisation of conditions

Ice thickness

Commercially available instruments such as the Vitrobot (Thermo Fisher Scientific) automate the plunge-freezing process. They allow for adjustment of several parameters to optimise plunge-freezing and yield frozen grids with suitable ice. Thick ice reduces image contrast and allows particles to overlap. Ice that is too thin excludes particles and may force their adsorption and denaturation at the air-water interface. An ideal grid would have uniformly thick ice across grid holes with a thickness approximately equal to the particle diameter plus 10 – 20 nm of ice separating particles from the air-water interface. In practice there is usually an ice thickness gradient

across the entire grid. Generally, regions of the grid that received less blotting have ice that is too thick for the electron beam to penetrate. The opposite end of the grid (where blotting was stronger) may contain areas with no ice. This is usually not a problem, as the ice thickness gradient across the grid often contains regions of suitably thin and vitreous ice from which data can be collected. A more important consideration is that ice thickness can also vary across individual grid squares and across holes. Care must then be taken to select appropriate grid holes for imaging. Data is usually not collected close to the grid mesh (around the inner edges of grid squares) as the ice at these locations tends to be thick enough to noticeably reduce contrast. An ice gradient over a hole may or may not be problematic depending on its steepness. In some cases, the ice thins towards the centres of holes enough to exclude particles. If the ice is too thick in certain areas, particles may overlap and even if they do not, contrast will be reduced. Parameters such as the blotting duration and blotting force allow users to optimise the plunge-freezing process and fairly routinely produce grids suitable for high resolution data collection.

Particle stability & distribution

The behaviour of particles during grid preparation can be influenced by the chosen plunge-freezing conditions. For example, long wait times, especially between blotting and plunge-freezing, lead to increased water evaporation. As water evaporates the concentration of solutes (e.g. buffer molecules, salts and detergents) increases. Some proteins may be sensitive to sudden large changes in pH or salt concentration. Particle behaviour is also affected by other variables such as protein solution concentration, choice of buffers and additives, type of EM grid and grid treatments. These should all be optimised to give uniformly distributed monodisperse particles that adopt random orientations over the grid holes. Achieving all these goals simultaneously is often challenging.

Ideally protein molecules would only reside over the holes of an EM grid, suspended in a thin ice layer. In practice specimens often preferentially adsorb to the support film resulting in a lower-than-expected particle concentration over the unsupported holes. This issue can sometimes be

alleviated by increasing protein solution concentration or by applying an additional continuous layer of carbon or graphene monolayer to the grid. Such layers provide a physical surface onto which particles can adsorb and will typically increase the apparent particle concentration over grid holes. The ideal protein concentration varies between grids with and without an additional support layer. Concentrations as low as 50 nM may be used when an additional support is present. This can be useful when working with low abundance proteins or proteins that do not concentrate well.

Particle orientation

Ideally, particles adopt random orientations and remain fully hydrated (i.e. distant from the air-water interface). Most specimens however adopt some degree of preferential orientation due to their interactions with the air-water interface or with the support film (if one is used). Presumably the surface properties of a molecule, such as the presence of hydrophobic patches, cause certain regions to interact more strongly with these interfaces. Cases of strongly preferred orientation hinder structure determination efforts and adsorption to the air-water interface can partially or completely denature proteins (Noble et al., 2018).

Many different methods exist to alleviate preferred orientation in cryoEM (Tan et al., 2017, Klebl et al., 2020 and D'Imprima et al., 2021). Their success is usually sample-dependent, meaning a trial-and-error approach must be adopted. Additives (primarily detergents) that alter the surface tension of water can prevent particles from adsorbing to the air-water interface, giving a less biased distribution of orientations (e.g. the detergent CHAPSO as used in Chen et al., 2019). The downside to this is that additives (such as detergents or glycerol) usually reduce particle contrast and may interfere with the structure of protein complexes. Screening many different additives at different concentrations in cryoEM can become time-consuming and resource intensive. A higher protein concentrations and different blotting parameters during plunge-freezing are often also required when detergents are used.

Another approach to addressing preferred orientation is to apply a support layer, e.g. a monolayer of hydrophilized graphene or graphene oxide, to the EM grid (Naydenova et al., 2019). Many proteins will preferentially adsorb to a physical support layer instead of the air-water interface. This can both reduce the level of preferential orientation (although not always) and prevent proteins denaturing at the air-water interface (D'Imprima et al., 2019).

If these approaches are unsuccessful, collecting data at a high stage tilt angle (20 – 40°) offers a sample-independent method of ameliorating the effects of preferred orientation. Collecting cryoEM images at high tilt angles introduces several complications that must be corrected for during image processing. Beam-induced movement of specimens increases with higher tilt angles, however this can be minimised by using gold grids (e.g. commercially available UltrAufoil or Au-flat grids). Tilting the specimen also introduces a defocus gradient across the image. Particles in the same image will be situated at different z-heights relative to each other. Image processing of tilted data can be performed using conventional methods. However, it is critical to account for the defocus gradient by determining the CTF parameters of individual particles. Many cryoEM software packages have implemented per-particle CTF estimation (such as CTF refinement in Relion or local refinement using Gctf). There is often a global loss in resolution in a tilted cryoEM data set. However, despite this resolution loss, tilting still improves map quality for specimens exhibiting strongly preferred orientations (Tan et al., 2017).

Current sample preparation methods have been identified as one of the major limitations in cryoEM (Glaeser, 2016). New methods are constantly under development and considerable effort has been put into making newer instruments fully automated. The next generation of grid preparation instruments will avoid the use of blotting and require sub-microlitre volumes of sample. These instruments, such as the Chameleon (discussed further in **Section 3.5**), also aim to minimise the time between sample application and freezing since this may reduce contact time

with the air-water interface and improve the orientation distribution of particles (Razinkov et al., 2016 and Arnold et al., 2017).

2.4. Principles of image formation and detection in the electron microscope

Electron microscopy is analogous to light microscopy. Both instruments comprise a radiation source, a series of lenses and a detector. The resolution of a microscope is limited to about half the wavelength of the radiation source used. Visible light has a wavelength of 400 – 700 nm which limits the resolving power of a light microscope to about 200 nm (2000 Å) (Penczek, 2010). Electron microscopes use high energy electrons instead of light. Electrons at an acceleration voltage of 300 keV have a wavelength of about 2 pm (0.02 Å) (Penczek, 2010). Therefore, the theoretical resolution is essentially unlimited for biological samples. In practice, properties of the electron microscope (and of the sample itself) limit the attainable resolution.

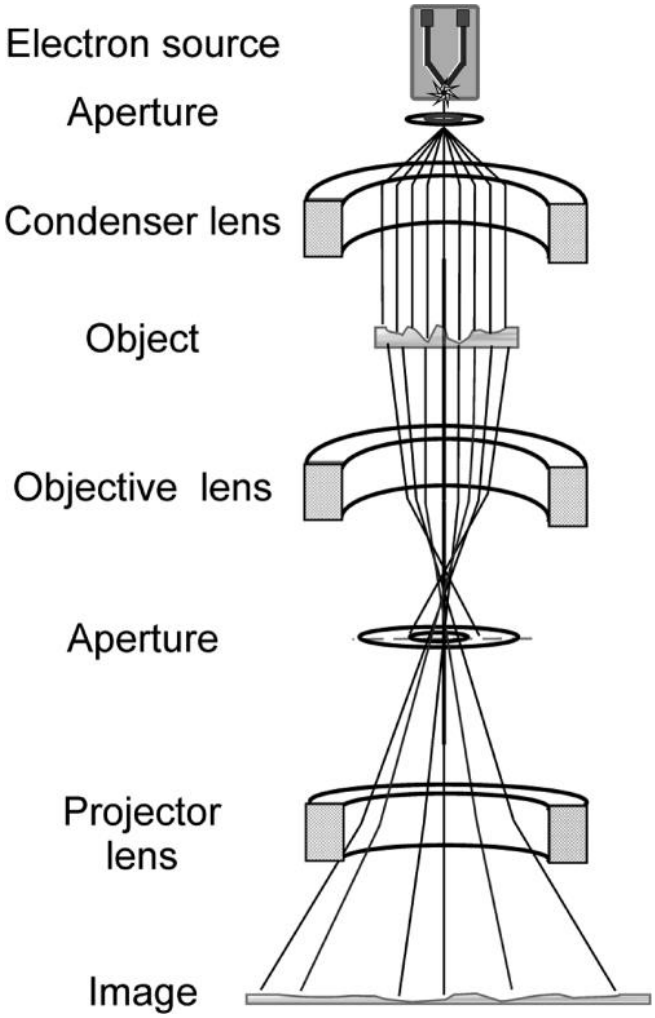

Figure 2-2. A simplified diagram of a transmission electron microscope. Reproduced from Orlova and Saibil (2011).

2.4.1. Electron sources

Electrons are emitted from an electron source at acceleration voltages of 80 - 300 keV. Several types of electron sources are commonly used. In low-end electron microscopes tungsten filaments and LaB_6 crystals are common. High-end microscopes use a field emission gun (FEG). FEGs use a smaller source size than other electron sources and generate a brighter, more coherent electron beam.

2.4.2. Lenses

A series of electromagnetic lenses manipulate electrons similarly to how optical lenses focus and direct visible light. The divergent electron beam is first focused into a parallel beam by the condenser lenses. A condenser aperture is used to improve contrast by restricting diverging rays. Together, the condenser lenses and choice of aperture control the brightness of the beam illuminating the sample. The objective lens is the most important optical component of an electron microscope. It focuses the beam to produce an image of the sample and provides the primary magnification (20 – 50x). The focus of the image is controlled by adjusting the strength of the objective lens. A series of intermediate and projector lenses further magnify the image onto a phosphorescent screen or a detector.

2.4.3. Interaction of the beam with the sample

Since biological samples consist of low atomic number elements, most (~80%) electrons pass through the sample without interacting. Electrons can interact with the sample destructively (inelastic scattering) or non-destructively (elastic scattering). Elastic scattering occurs when electrons are scattered by atomic nuclei without changing energy. Approximately 5% of the electron beam will interact with the sample in this way. Inelastic scattering accounts for ~15% of interactions and occurs when an electron transfers energy to the specimen. Energy transfer to the specimen results in radiation damage by ionisation, X-ray emission, generation of free-radicals or chemical bond rearrangement. Secondary electrons may also be generated and cause further damage to the specimen. To limit radiation damage during imaging samples are imaged under "low dose" conditions. An exposure of 10 e^-/A^2 or less will already begin to degrade high resolution features. This dictates the fundamental limit to resolution in cryoEM.

2.4.4. Contrast transfer

The contrast in an image represents its intensity variations. Objects produce contrast in two ways. Amplitude contrast is produced when the specimen absorbs part of the electron beam. Phase contrast is produced when an object deflects part of the beam resulting in a phase shift between the scattered and unscattered electrons. Biological specimens are electron transparent due to their chemical composition and do not generate significant amplitude contrast (values of 5 – 10% are commonly cited). Elastically scattered electrons have an altered path through the specimen that results in a phase shift compared to the unscattered beam. This small angle phase shift produces phase contrast and contributes to image formation in transmission EM. Objects in cryoEM are therefore described as *weak phase objects* and their contrast is linearly related to their projected potential. This means that images of objects in cryoEM are projection images, analogous to a medical X-ray.

Since unstained biological samples scatter only around 5% of electrons and only at small angles, an in-focus image of a thin biological specimen will have essentially no contrast (i.e. it will be invisible). Contrast of weak phase objects can be increased during image acquisition by applying an additional phase shift to the scattered electron wave. This is achieved by deliberately under-focussing the objective lens.

2.4.5. Contrast transfer function

The contrast transfer function (CTF) is a mathematical function that describes how contrast is transferred to an image in terms of spatial frequency. It is an oscillatory function that crosses through zero at multiple points. Example CTF curves are shown in **Fig. 2-3A** and **Fig. 2-3B**. Some Fourier components are transferred with positive contrast while others have negative contrast. At points where the function crosses zero, no contrast is transferred to the image leading to a loss of information at certain spatial frequencies. Certain properties of the electron microscope

such as acceleration voltage and spherical aberration affect the CTF, but these are known and constants for a particular microscope setup. The level of defocus also affects the CTF but varies between images due to lens settings as well as sample height. To restore the image, the CTF of each micrograph (or particle) must be accurately determined. In addition, to restore the full level of detail of objects it is necessary to collect images through a range of different focus levels, typically ~ 0.5 – 3.0 μm under focus. This acts to "fill in the gaps" of the overall CTF of the object (**Fig. 2-3C** and **Fig. 2-3D**). Higher defocus values generate more contrast at low frequencies making particles easier to see (**Fig. 2-3B**, red curve). However, this comes at the cost of a stronger envelope function (not shown in **Fig. 2-3**). The envelope function is a result of imperfect imaging and can strongly attenuate high frequency (high resolution) signal.

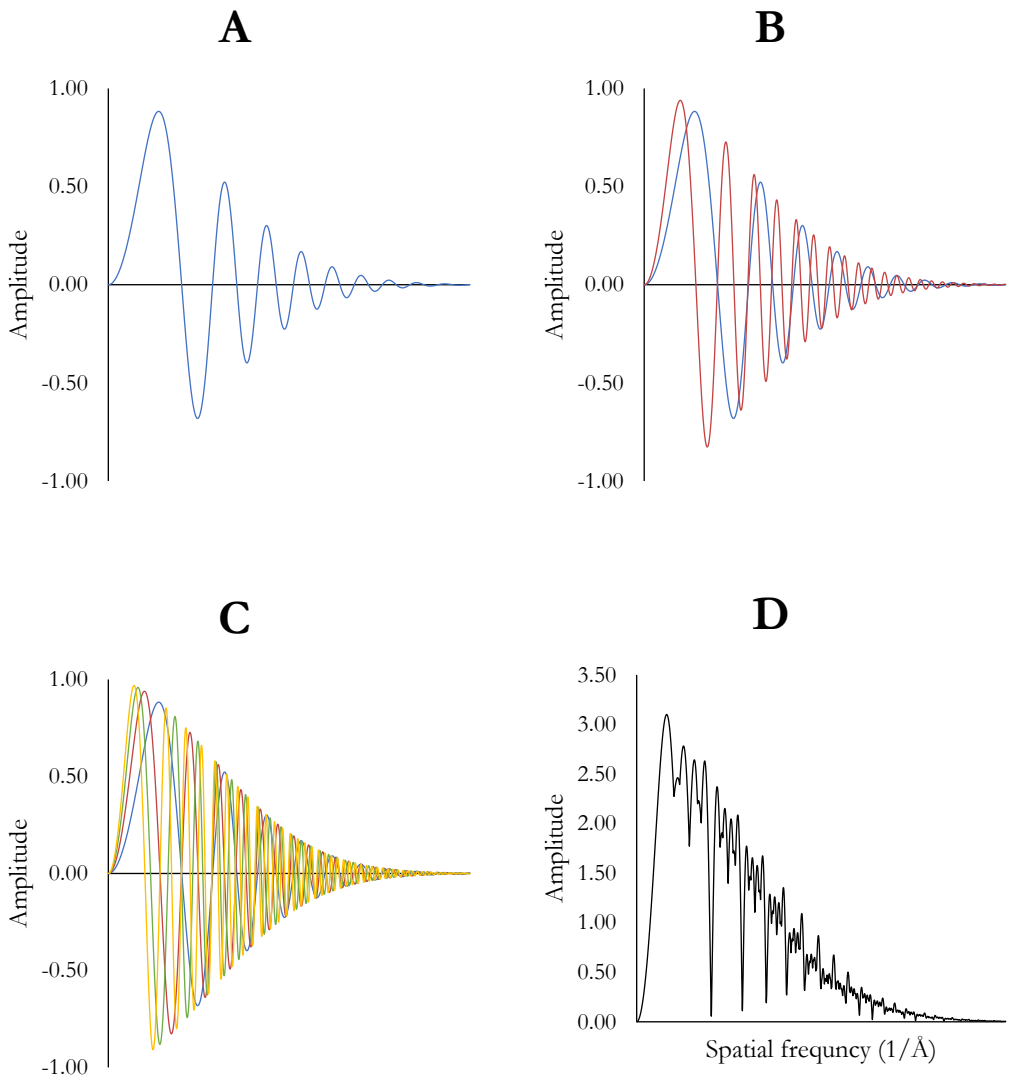

Figure 2-3. Simulated CTF curves. (**A**) Simulated CTF curve for a defocus value of 0.5 μm (blue). (**B**) Simulated CTF curves for defocus values of 0.5 μm (blue) and 1.0 μm (red). The red curve corresponds to an image further from focus. Contrast is increased at low spatial frequencies but the CTF oscillates more rapidly. The envelope function is the same in these simulated CTF curves, but would be stronger for the CTF at higher defocus in practice. (**C**) Overlaid simulated CTF curves for defocus values of 0.5 μm (blue), 1.0 μm (red), 1.5 μm (green) and 2.0 μm (yellow). The additional CTF curves begin to add contrast where there is none in previous plots. (**D**) During image processing, negative contrast values are phase-flipped, making them positive. Plot D shows the sum of the phase-flipped amplitude values for all four CTF curves in plot C. This shows the total contrast transferred to the image as a function of spatial frequency. Note that high spatial

frequency information is transferred less efficiently than low spatial frequency information. In all plots the x-axis is spatial frequency (1/Å) and extends to 0.5 $Å^{-1}$, corresponding to a real-space resolution of 2 Å. CTF curves were generated with e2ctfsim.py in EMAN version 2.2 with the following parameters set, acceleration voltage of 300 keV and a spherical aberration of 2.7 mm (Tang et al., 2007).

2.4.6. Detectors

Historically, photographic film was used to detect electrons. Following exposure each film had to be chemically treated, dried and digitised in a time-consuming and laborious process. Digital detectors such as charge-coupled devices (CCDs) replaced film as the detector of choice due to their ease of use and the added ability to automate data collection. Further advances now allow for the direct detection of electrons and several such direct electron detectors are available commercially. The higher readout speed of direct electron detectors (compared to CCD detectors) allows for the collection of movie data at speeds up to 1,500 frames per second. This allows individual movie frames to be aligned to correct for the slight movements of particles in the ice during data collection. The latest generation of direct detectors have high Detective Quantum Efficiency (DQE). DQE is a measure of how much the noise, or the mechanism of signal conversion degrades the original image (McMullan et al., 2016). A perfect detector would not add any noise and have a DQE of 1.0 (Bai et al., 2015). At Nyquist frequency, the DQE of the Gatan K3 camera is reported to be 0.35 – 0.4. The advent of direct electron detectors coupled with improvements in image processing software underpins what has been called the *Resolution Revolution* in cryoEM (Kuhlbrandt, 2014). Many recent reviews have highlighted the importance of these advances (Bai et al., 2015, Glaeser, 2019, Nogales, 2015 and Murata and Wolf, 2018).

2.5. Image processing

2.5.1. Motion correction

Direct electron detectors record a set of sub-frames (movies) instead of a single image. The individual sub-frames can be aligned and summed to minimise the effects of beam-induced motion. Data recorded by direct electron detectors is also dose fractionated. In early frames, particles have accumulated a lower total electron exposure than in later frames. This gives users the ability to discard or down-weight sub-optimal frames. Programmes for motion correction can down-weight later sub-frames as these have attenuated high resolution information. However, programmes will still typically use the full set of summed frames for CTF estimation as this gives the highest SNR over all spatial frequencies. Each sub frame is then weighted according to its dose based on expectations from analysis of radiation damage. Dose weighted summed images are then used for high resolution structure determination.

2.5.2. CTF Estimation

As discussed, cryoEM images are modulated by the CTF. Part of the CTF is defined by certain microscope settings (acceleration voltage and spherical aberration). The unknown part of the CTF (i.e. the defocus) must be determined for each experimental image. This is accomplished by fitting a theoretical model (similar to the one shown in **Fig. 2-3A**) to the power spectra of each micrograph. Once the CTF has been determined it can be used to correct the image. Initial CTF estimation is performed on a per-micrograph basis. At later stages the CTF can be refined on a per-particle basis through a different method, leading to more accurate CTF estimation and higher resolution reconstructions.

2.5.3. Particle picking and extraction

Particle picking is the process of selecting suitable particles from micrographs. Picking can be performed manually, but this is tedious and time-consuming to point of being impractical, especially when datasets contain thousands of micrographs and millions of particles. Most single-particle cryoEM software packages include automated or semi-automated particle picking programmes and it is sometimes necessary to trial several for optimal results. Automated particle picking is rarely perfect; algorithms frequently select high contrast areas such as ice crystals and edges of carbon. Some programmes have implemented machine learning approaches that must be trained on the dataset prior to actual picking. Once picked, particles are extracted from micrographs as individual images. Usually an image size (or box size) of 2.5x the particle diameter is sufficient for initial processing. Extracted particles are then ready for classification and alignment.

2.5.4. 2D Classification

Particle images are first analysed and clustered into homogenous groups, a process commonly called 2D classification. Analysis of the 2D images provides an initial assessment of the overall dataset quality.

Relion uses a maximum-likelihood approach for 2D classification. The user provides only the number of initial references (K) and the programme proceeds in an unsupervised manner. Particles are divided into *K* random subsets of equal size. The averaged image of each random subset is then used as a reference for alignment. Each particle image is compared to each reference in every possible orientation and a number of translations. Particle images are then assigned probability weights for each orientation and each reference. A weighted average of all possible class assignments is then used to calculate a 2D class average. These class averages are then used as the

references for the next iteration of 2D classification. In practice this process usually converges so that each particle image contributes to a single 2D class average.

2D classification of high-resolution data usually produces a mixture of "good" and "bad" classes. Good classes show well-resolved, recognisable features of proteins and symmetry may be apparent. Bad classes often represent damaged or misaligned protein, ice contaminants or edges of the support film. Particle images from good classes can be subjected to additional rounds of 2D classification or further processing.

2.5.5. 3D refinement

The goal of 3D refinement is to assign accurate orientations (three Euler angles; α, β, γ) and positions (two translations; x, y) to each experimental particle image relative to an initial 3D reference. Before refinement can proceed an initial map (commonly called initial model) must be generated to act as the reference. The initial map is an estimate of the final 3D map. It can be derived experimentally, computed from the data or generated from existing cryoEM or crystal structures, usually filtered to low resolution to avoid model bias. During refinement, reference images are generated by projecting the initial map in multiple different directions and comparing them to each particle image. This procedure is called projection matching and variations of it are used by many 3D refinement programmes. In Relion, particle images are compared with each reference projection and assigned a probability using maximum likelihood. This approach avoids assigning discrete orientations which can result in the procedure falling into local minima. The reference images provide probability-weighted angular assignments for the experimental particle images. After all particle images have been evaluated, a 3D reconstruction is computed and used to generate a new set of reference projections for the next iteration of 3D refinement. Refinement continues until the angular assignments of particles converge.

In Relion (and other image processing programmes), the data set is split into two equal halves. Usually this is done by assigning even numbered particles to one half and odd numbered particles to the other half. Each half is then refined separately through iterative refinement. Progress is monitored at the end of each iteration of 3D refinement using the Fourier shell correlation (FSC). The FSC gives information on the signal-to-noise ratio (SNR) as a function of spatial frequency. An FSC curve is produced by comparing the 3D reconstructions derived from each half of the data. Correlation coefficients for each resolution shell are computed from the Fourier transforms of each half map and plotted. This approach reduces the likelihood of aligning noise to references (i.e. overfitting) and is known as the *gold-standard FSC* (Scheres and Chen, 2012).

Map resolution and quality can be evaluated from the gold-standard FSC curve. The most common way to report map resolution is to read out the spatial frequency value (the reciprocal of resolution) when the FSC is equal to 0.143. This value is based on a comparison between EM maps and crystal structures (Rosenthal and Henderson, 2003). The reported resolution of a cryoEM density map represents an average over the entire structure. Some regions may be well resolved with visible amino acid sidechains and secondary structure. Less well-resolved regions hint at heterogeneity in the data and local resolution maps are often computed for cryoEM data to highlight these regions.

2.5.6. 3D classification

Structural heterogeneity is present in all cryoEM datasets. Populations of particles may have subtle differences in structure that represent different protein conformations or the presence of additional density from subunits, substrates or ligands. In "consensus" refinements (refinements that use all the particles of a dataset following 2D classification) these details may only be visible at low SNR levels, or not at all. To resolve structural heterogeneity, particles must be sorted into more homogenous groups using 3D classification.

In 3D classification, multiple 3D reconstructions are simultaneously refined. The user provides the number of classes and a single 3D reference used to initialise each of the 3D refinements. Particles are randomly assigned to 3D classes and compared to projections generated from each reference map. Each particle is assigned a probability for each orientation, translation and 3D class. Once all particles have been assigned these probabilities, new 3D reference maps are computed for the next iteration. Classification then proceeds in an iterative manner for a set number of cycles. The use of multiple different classes prevents the use of gold-standard FSC in 3D classification. Therefore, after 3D classification has converged and a suitable class is identified, particles from that class can be refined using the gold-standard FSC.

The number of different structures present in the data is usually unknown so that multiple runs of 3D classification with different numbers of classes should be performed to generate useful results. As in 2D classification, the first run of 3D classification typically produces one or more "junk" classes, these should be examined and discarded. The remaining particles can then subjected to further classification strategies.

2.5.7. 3D classification without alignments

In Relion, 3D classification can be used in several different ways. In a "default setting" 3D classification run, both alignment and classification of particles are performed during each iteration. However, when focussing on small, poorly resolved regions of interest in particles using a mask, the SNR is usually too low to accurately align particles. In these situations, alignments can be skipped so that only classification is performed on the images during each iteration. One caveat of this is that particles must already have alignments assigned, for example from a previous consensus refinement.

2.5.8. Post-processing

Post-processing, as implemented in Relion, has two goals, i) to calculate a higher resolution FSC and ii) to sharpen the map. The two half maps generated during each iteration of 3D refinement are not masked when calculating FSC curves. Because the half maps are unmasked, noise from the solvent outside of the structure is included in the 3D reconstructions and the resolution estimations. Inclusion of noise lowers the FSC due to poorer correlations. Solvent-derived noise can be removed by applying a mask to the structure. Masks are 3D volumes that multiply areas inside the mask by 1 and areas outside by 0. Therefore, by masking the individual half maps prior to FSC curve calculation, a more accurate measure of resolution can be determined. The solvent mask introduces convolution effects, which are corrected for during postprocessing (Scheres, 2016). High resolution data is attenuated during image formation, detection, and processing. This can give reconstructions a smooth and featureless appearance. To restore the high-resolution features of a map, it is sharpened by applying a negative B-factor to down weight the over-represented low resolution components, making high resolution features more visible in the final cryoEM reconstruction.

2.5.9. CTF refinement and Bayesian polishing

During initial image processing, CTF estimation is performed on a per-micrograph basis using programmes such as CTFFIND or Gctf. This provides the same average defocus value for all particles in a single micrograph. In the latest versions of Relion, CTF parameters (defocus and astigmatism) can be refined with the use of a 3D reference structure (Zivanov, 2018). This approach produces more accurate results and typically leads to modest increases in attained resolution (anywhere from 0.5 – 1.5 Å in our experience with GroEL complexes).

Similarly, initial motion correction is performed on a per-micrograph basis by splitting the micrograph into tiles (typically 6x5 tiles) and determining the motion of each tile. Bayesian

polishing estimates the motion trajectories of individual particles, and the amount of cumulative radiation damage (Zivanov, 2019). Radiation damage cannot be removed from the data, but its effect can be reduced by downweighing the later, high-dose movie frames.

Chapter III

3. Methods

3.1. Protein expression and purification

3.1.1. Escherichia coli GroEL

Plasmid pTrcESL, a kind gift from Dr. Peter Lund (University of Birmingham, UK), was transformed into BL21 *E. coli*. Cells were grown in LB media at 37 °C using baffled flasks. Overexpression of GroELS was induced by addition of 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG) at an A_{600} of 0.6 – 0.8. After 3.5 hours cells were harvested by centrifugation, resuspended in chilled lysis buffer (50 mM Tris, pH 7.4), and disrupted using an EmulsiFlex-C3 (Avestin, Ottawa, Canada). Cellular debris and insoluble material were removed by centrifugation at 55,000 RCF at 4 °C for 1 hour. Subsequent purification and dialysis steps were performed at 4 °C unless stated. Soluble lysate was filtered through a 0.45 μm syringe filter and loaded onto a 140 mL Q Sepharose FastFlow ion-exchange column (GE Healthcare) equilibrated in buffer A (50 mM Tris, pH 7.4, 1 mM DTT). The column was washed with 20% buffer B (50 mM Tris, pH 7.4, 1 mM DTT, 1M NaCl) until A_{280} reached a steady baseline. Protein was eluted with a 20 – 50% 15 column volume gradient of buffer B and fractions were collected. Elution of GroEL was confirmed by SDS-PAGE and fractions corresponding to the latter half of the GroEL peak were pooled. Solid ammonium sulphate was added slowly to the pooled fractions to a final concentration of 1.2 M and stirred overnight. Pooled fractions from ion-exchange chromatography were filtered through a 5 μm syringe filter and loaded onto a 22 mL Source 15ISO hydrophobic interaction column (GE Healthcare) equilibrated in buffer C (50 mM Tris, pH 7.4, 1 mM DTT, 1.2 M $NH_4(SO_4)_2$. The column was washed with buffer C until the A280 reached a

steady baseline < 50 mAU. GroEL was eluted with a 14 column volume reverse-gradient of 100 – 0% buffer C to buffer A. Peak fractions were analysed by SDS-PAGE. The protein concentration and tryptophan fluorescence of each peak fraction were measured. GroEL has no tryptophan residues therefore tryptophan fluorescence is an indicator of contamination by bound Trp-containing proteins. Tryptophan fluorescence was measured using a FluoroMax 3 (Horiba). Protein concentration was measured by Pierce BSA protein assay. Fractions with high protein concentration and low fluorescence were pooled and dialysed overnight against 4 L of buffer D (50 mM Tris, pH 7.4, 50 mM KCl, 1 mM DTT). Usually the final peak fractions from hydrophobic interaction chromatography consistently yielded the purest GroEL. Dialysed protein was concentrated to 20 mg/mL and 2 mL of buffer E (1:1 mixture of buffer D and Affi-gel Blue resin) was added per 30 mg of total protein. Methanol was added dropwise to 20% v/v while stirring and the sample was rocked gently at room temperature (~ 21 °C) for 2 hours. GroEL was recovered in the supernatant by filtration through a 1.2 μm syringe filter and the resin was washed once with its volume of buffer D. Eluates were pooled and dialysed overnight against 4 L of buffer D. Purified GroEL was filter-sterilised and concentrated to 30 – 40 mg/mL. Aliquots of concentrated GroEL were flash frozen in liquid nitrogen and stored at -80 °C until use.

3.1.2. Escherichia coli GroES

Plasmid pTrcESL was transformed into BL21 *E. coli*. Cells were grown in LB media at 37 °C using baffled flasks. Overexpression of GroELS was induced by addition of 0.5 mM IPTG at an A_{600} of 0.6 – 0.8. After 3.5 hours cells were harvested by centrifugation, resuspended in chilled lysis buffer (50 mM Tris, pH 7.4), and disrupted using an EmulsiFlex-C3 (Avestin, Ottawa, Canada). Cellular debris and insoluble material were removed by centrifugation at 55,000 RCF at 4 °C for 1 hour. Soluble lysate was transferred to a fresh tube and acidified to pH ~ 5.0 by dropwise addition of 1 M acetic acid. Precipitated material was removed by centrifugation at 25,000 RCF at 4 °C for 15 min. Subsequent purification and dialysis steps were performed at 4 °C unless stated. Acidified

soluble lysate was filtered through a 0.45 μm syringe filter and loaded onto a 125 mL SP Sepharose FastFlow ion-exchange column (GE Healthcare) equilibrated in buffer A (50 mM NaOAc, pH 5.0). The column was washed with buffer A until A_{280} reached a steady baseline. Protein was eluted with a 0 – 50% 12.5 column volume gradient of buffer B (50 mM NaOAc, pH 5.0, 1 M NaCl) and fractions were collected. Elution of GroES was confirmed by SDS-PAGE. Fractions that corresponded to the GroES peak were pooled, neutralised to pH 7.5 – 8.0 with Tris-base, and dialysed overnight against 4 L of buffer C (50 mM Tris, pH 7.4). Dialysed protein was loaded onto a 24 mL Source 15Q ion-exchange column equilibrated in buffer C. The column was washed with buffer C until the A280 reached a steady baseline < 50 mAU. GroES was eluted with a 0 – 50% 20 CV gradient of buffer D (50 mM Tris, pH 7.4, 1 M NaCl). Elution of GroES was confirmed by SDS-PAGE. Fractions that corresponded to GroES were pooled and dialysed overnight against 4 L buffer C. Purified GroES was filter-sterilised and concentrated to ~ 15 mg/mL. Aliquots of concentrated GroES were flash frozen in liquid nitrogen and stored at -80 °C until use.

3.1.3. Rhodospirillum rubrum RuBisCO

Plasmid pT7Ru, a kind gift from Dr. Wayne Fenton (Yale University, USA), was transformed into BL21 (DE3) *E. coli*. Cells were grown in LB media in baffled flasks at 20 °C for 24 hours in the absence of induction. Cells were harvested by centrifugation, resuspended in lysis buffer (50 mM Tris, pH 7.4), and disrupted using an EmulsiFlex-C3 (Avestin, Ottawa, Canada). Cellular debris and insoluble components were removed by centrifugation at 55,000 RCF at 4 °C for 1 hour. Subsequent purification and dialysis steps were performed at 4 °C unless stated. Soluble lysate was filtered through a 0.45 μm syringe filter and loaded onto a 140 mL Q Sepharose FastFlow ion-exchange column (GE Healthcare) equilibrated in buffer A (50 mM Tris, pH 7.4, 1 mM DTT). The column was washed with buffer A (50 mM Tris, pH 7.4, 1 mM DTT) until A_{280} reached a steady baseline. Rubisco was eluted with a 0 – 50% 10 column volume gradient of buffer B (50 mM Tris, pH 7.4, 1 mM DTT, 1M NaCl) and fractions were collected. Elution of Rubisco was

confirmed by SDS-PAGE and peak fractions were pooled and dialysed overnight against 4L of buffer C (20 mM Tris, pH 7.4, 1 mM DTT). Pooled fractions from ion-exchange chromatography were filtered through a 1.2 μm syringe filter and loaded onto a 60 mL Affi-Gel Blue column equilibrated in buffer C. The column was washed with buffer C until the A_{280} reached a steady baseline. Rubisco was eluted with a 0 – 50% 4 column volume gradient of buffer D (20 mM Tris, pH 7.4, 1 mM DTT, 1M NaCl). Peak fractions were analysed by SDS-PAGE, pooled and dialysed overnight against 4 L of buffer A. Pooled fractions from dye-ligand chromatography were filtered through a 1.2 μm syringe filter and loaded onto a 24 mL Source 15Q column equilibrated in buffer A. The column was washed with buffer A until the A_{280} reached a steady baseline. Rubisco was eluted with a 0 – 50% 10 column volume gradient of buffer B. Peak fractions were analysed by SDS-PAGE and the purest fractions were pooled and dialysed overnight against 4 L of buffer A. Purified Rubisco was concentrated to 10 – 20 mg/mL and filter-sterilised. Aliquots of concentrated Rubisco were flash frozen in liquid nitrogen and stored at -80 °C.

3.2. Formation of GroEL-substrate complexes

Rubisco was unfolded in unfolding buffer (50 mM HEPES-KOH, pH 7.5, 8 M urea) at 21 °C for at least 30 minutes. Binary complexes of GroEL bound to non-native Rubisco were prepared by five successive 100-fold dilution steps of denatured Rubisco into GroEL-containing HKM buffer (50 mM HEPES-KOH, pH 7.5, 10 mM KOAc, 10 mM $Mg(OAc)_2$, 2 mM DTT, 1 μM GroEL tetradecamer). The final concentration of Rubisco monomers was 4 μM, giving a 4-fold molar excess of Rubisco monomers over GroEL oligomers. After the final addition of non-native Rubisco, the sample was incubated at 21 °C for 10 minutes with periodic gentle mixing via gentle pipetting. Complexes were centrifuged at 16,200 RCF at 21 °C for 10 minutes to pellet aggregated protein. The presence of binary complexes was confirmed by BN-PAGE and/or native mass spectrometry. Binary complexes were freshly prepared for all experiments.

3.2.1. Formation of nucleotide-bound complexes

Binary complexes were prepared as described in **Section 3.2**. ADP or ATP were added to a final concentration of 3 mM. After incubation at 21 °C, KF and $BeSO_4$ were added to a final concentrations of 20 mM KF and 2 mM $BeSO_4$, to form the ATP ground state analogue ADP·BeF_x. To form the ATP transition state analogue ADP·AlF_x, $BeSO_4$ was replaced by $KAl(SO_4)_2$ at the same concentration.

3.2.2. Formation of GroES-bound complexes

Binary complexes were prepared as described in **Section 3.2**. The concentration of GroEL-substrate complexes was measured by Pierce BSA protein assay (ThermoFisher). GroES was then added at the same concentration. For the Chameleon cryoEM data set of GroEL-GroES-Rubisco-ADP·AlF_x, the measured concentration of GroEL-Rubisco was ~7 μM, we therefore added 7 μM GroES.

3.3. Attempts to improve sample homogeneity

BN-PAGE showed that GroEL-Rubisco samples were heterogeneous, containing both apo-GroEL and at least two species of GroEL-Rubisco. While heterogeneity in cryoEM samples can be addressed during image processing, it is generally advantageous to produce a homogeneous sample (Scheres, 2016).

Two approaches were explored to attempt to produce a more biochemically homogenous sample of GroEL-Rubisco. The first was to separate apo-GroEL from GroEL-Rubisco using chromatographic techniques. Biochemical separation of substrate-bound GroEL from apo-GroEL is challenging due to the near identical physical and chemical properties of the two complexes. Protocols to purify GroEL may include a hydrophobic interaction chromatography (HIC) step, allowing isolation of GroEL with minimal contaminants. In this step, GroEL elutes from the HIC column in two stages, represented by two UV absorbance (A_{280}) peaks. Each peak

often has a prominent shoulder. The shoulder of the second peak (i.e. the last-eluting protein from the column) consistently contained the highest-purity GroEL, as judged by tryptophan fluorescence and SDS-PAGE. The presence of bound non-native protein must affect the physical properties of the complex enough to allow chromatographic separation. Despite this, efforts to purify GroEL-Rubisco from the heterogenous mixture using either hydrophobic interaction or high-resolution ion-exchange chromatography proved unsuccessful.

The second approach to obtain a more homogenous sample was to increase the occupancy of bound non-native Rubisco. Various conditions aimed at achieving this were trialled and occupancy of non-native Rubisco was judged by BN-PAGE. Increasing the molar excess of non-native Rubisco from 2 to 4 appeared to increase the occupancy of bound substrate. However, higher molar excesses had no observable effect when judged by BN-PAGE. There was also no observable difference in occupancy when the binary complexes were formed at 4 °C or at 21 °C.

3.4. Negative stain EM

CF-400 grids with an amorphous carbon support layer were glow discharged at 30 mA for 60 seconds using a Pelco easiGlow system (Ted Pella, Inc). Three microlitres of sample were applied to grids and allowed to adsorb for 20 seconds. Excess solution was removing by blotting and the top face of the grid was touched to a 50 μL drop of 2% uranyl acetate (UA) on parafilm then blotted again. The grid was touched against a second drop of UA and held against it for 20 seconds, excess solution was again blotted. The grid was left to air dry for at least 15 minutes prior to imaging. Negative stain images were collected on a Tecnai T12 operated at 120 keV. Images were recorded at a nominal magnification of 52,000 on a 4k x 4k Gatan CCD camera. Negative stain images were processed up to 2D classification using a typical workflow in EMAN version 2.2.

3.5. CryoEM grid preparation – The Vitrobot and Chameleon

For most cryoEM experiments, grids are prepared using a Vitrobot (Thermofisher). A Vitrobot uses a vertical pair of tweezers to hold an EM grid inside a humidity-controlled chamber. Sample is applied to one side of the grid via manual pipetting. When a Vitrobot cycle starts, two pads with blotting paper are touched against both sides of the grid for a set time. After this time, usually several seconds, the pads are withdrawn, and the grid is plunged into liquid ethane cooled by liquid nitrogen. The thin layer of buffer in which the sample resides is vitrified. CryoEM grids typically require some optimisation by experimentation with different Vitrobot parameters such as the blot time and blot force. For an experienced user it takes several seconds to pipette a sample onto a grid, withdraw the pipette, and start the Vitrobot process. A typical blot time is 3 – 5 s. Overall, the total time between sample application and vitrification is around 5 – 10 s. During this time, the sample may interact with the air-water interface 10^3 to 10^4 times (Klebl et al., 2020). These interactions can lead to adsorption of proteins to the air-water interface. This can cause some proteins to denature partially or completely. The presence of non-native substrate bound to GroEL causes the whole complex to adsorb to the air-water interface in a strongly preferred orientation.

The Chameleon is a fully automated vitrification instrument. Chameleon is the commercially available version of "Spotiton", an instrument developed in an academic environment over the last decade (Dandey et al., 2018). It reduces the time between sample application and plunge-freezing. Chameleon uses a piezo electric dispenser to deliver pL droplets of sample onto a self-wicking nanowire grid milliseconds before the grid is plunged into liquid ethane. Because the Chameleon is a blot-free instrument, self-blotting (or self-wicking) grids are used. The self-blotting grids have a mesh size of 300 and hole size and spacing of 1.2 and 0.8 μm. The grid bars of these self-wicking grids are coated with nanowires that effectively act as blotting paper, reducing the liquid to a thin layer suitable for cryoEM data collection (Wei et al., 2018). These grids also have a

holey carbon film, similar to commonly used grids for cryoEM. Sample volume requirements are low and 5 µL is enough to process ~100 cryoEM grids. This allows low abundance samples to be concentrated. Dispense-to-freeze times can be as low as 54 ms, but are typically in the range of 100 – 200 ms. The data we collected came from grids with a dispense-to-freeze time of 54 – 1000 ms. We found the Chameleon to be a useful instrument in reducing the interaction of GroEL-substrate complexes with the air-water interface. This generally led to an increase in more "useful" side views of GroEL. In our cryoEM data sets, we did not notice a significant different between the faster and slower dispense-to-freeze times. Initial reconstructions from "fast" and "slow" grids were similar, as were orientation distribution plots. The reduction from 5 – 10 s (using a Vitrobot), to under 1 s using Chameleon was enough to yield significant improvements. Further improvements would only likely be gained by reducing this time further to single-digit milliseconds or lower. Microfluidic spraying devices boasting dispense-to-freeze times as low as 6 ms are currently also being developed (Kontziampasis et al., 2019).

3.6. CryoEM

For this thesis, a total of eleven cryoEM data sets were collected. These were: (**1**) apo GroEL, (**2**) GroEL-Rubisco (untilted data set), (**3**) GroEL-Rubisco (tilt data set), (**4** and **5**) GroEL-Rubisco (two data sets from Chameleon grids), (**6**) GroEL-Rubisco-ATP, (**7**) GroEL-Rubisco-ADP·BeF_x (tilt data set), (**8**) GroEL-Rubisco-ADP·AlF_x (tilt data set), (**9** and **10**) GroEL-Rubisco-ADP·BeF_x (two data sets from Chameleon grids) and (**11**) GroEL-GroES-Rubisco-ADP·AlF_x (Chameleon grid). Data sets with identical cryoEM collection parameters and grid preparation methods are grouped together to avoid repetition. Data collection parameters for all cryoEM data sets is summarised in **Appendix 1**.

3.6.1. Apo GroEL

C-Flat R2/2 grids were glow discharged at 30 mA for 60 seconds using a Pelco easiGlow system. Three microlitres of Apo GroEL (1.5 μM) were applied to grids, blotted and plunged into liquid ethane cooled by liquid nitrogen. Plunge-freezing was performed using a Vitrobot mark IV (Thermo Fisher Scientific) operating at 100% humidity and 4 °C. Blot time was set to 3.5 s and blot force set to -10. Movies (50 frames) were collected using the EPU software on a Titan Krios transmission electron microscope (Thermo Fisher Scientific) operating at 300 keV, equipped with a Gatan K2 Summit direct electron detector and Gatan energy filter. The defocus range was set between -1 and -2.7 μm, and the total exposure was 39.2 electrons/Å^2. The pixel size was 1.048 Å/pixel.

3.6.2. GroEL-Rubisco (untilted)

C-Flat R2/2 grids were glow discharged at 30 mA for 60 seconds using a Pelco easiGlow system. Three microlitres of Apo GroEL (1.5 μM) or GroEL-Rubisco (0.8 μM) were applied to grids, blotted and plunged into liquid ethane cooled by liquid nitrogen. Plunge-freezing was performed using a Vitrobot mark IV (Thermo Fisher Scientific) operating at 100% humidity and 4 °C. Blot time was set to 4 s and blot force set to -10. Movies (40 frames) were collected using the EPU software on a Titan Krios transmission electron microscope (Thermo Fisher Scientific) operating at 300 keV, equipped with a Gatan K2 Summit direct electron detector and Gatan energy filter. The defocus range was set between -1 and -2.7 μm, and the total exposure was 39.2 electrons/Å^2. The pixel size was 1.048 Å/pixel.

3.6.3. GroEL-Rubisco (tilt data set)

C-Flat R2.0/2.0 grids were glow discharged at 30 mA for 60 seconds using a Pelco easiGlow system. A monolayer of graphene oxide was added as previously described (Cheng et al., 2020). Three microlitres of GroEL-Rubisco (diluted to 55 nM) were applied to grids, blotted and plunged

into liquid ethane cooled by liquid nitrogen. Plunge-freezing was performed using a Vitrobot mark IV (Thermo Fisher Scientific) operating at 100% humidity and 4 °C. Blot time was set to 10.5 s, blot force set to -10, the wait time between blotting and plunge-freezing was set to 10 s. Movies (40 frames) were collected using the EPU software on a Titan Krios transmission electron microscope (Thermo Fisher Scientific) operating at 300 keV, equipped with a Gatan K2 Summit direct electron detector in counting mode and Gatan energy filter. The defocus range was set between -1.5 and -2.7 μm, and the total exposure was 39.2 electrons/Å^2. The pixel size was 1.048 Å/pixel. A stage tilt of 30° was set at the start of image acquisition. This was increased to 40° after collecting ~3500 movies.

3.6.4. GroEL-Rubisco (two data sets from Chameleon grids)

Grids were prepared using a Chameleon instrument (SPT Labtech) operated by Michele Darrow. GroEL-Rubisco was concentrated to 3.4 μM. Movies (48 frames) were collected using the EPU software on a Titan Krios transmission electron microscope (Thermo Fisher Scientific) operating at 300 keV, equipped with a Gatan K2 Summit direct electron detector in counting mode and Gatan energy filter. The defocus range was set between -1.4 and -3.0 μm, and the total exposure was 40.2 electrons/Å^2. Due to the low number of particles, we collected images at a lower magnification at a pixel size of 1.34 Å/pixel.

3.6.5. GroEL-Rubisco-ATP (tilt data set)

UltrAuFoil R2/2 grids were glow discharged at 30 mA for 60 seconds using a Pelco easiGlow system. Three microlitres of GroEL-Rubisco (1 μM) were applied to grids. ATP (3 mM) was added, then grids were immediately blotted and plunged into liquid ethane cooled by liquid nitrogen. Plunge-freezing was performed using a Vitrobot mark IV (Thermo Fisher Scientific) operating at 100% humidity and 4 °C. Blot time was set to 5 s, blot force set to -10. Movies (50 frames) were collected using the EPU software on a Titan Krios transmission electron microscope

(Thermo Fisher Scientific) operating at 300 keV, equipped with a Gatan K3 direct electron detector operating in super-resolution mode and Gatan energy filter. Data set was collected using Krios III at eBIC (Clare et al., 2017). The defocus range was set between -1.5 and -2.7 μm, and the total exposure was 50 electrons/Å^2. The pixel size was 0.53 Å/pixel in the super-resolution movies. A stage tilt of 35° was set at the start of image acquisition.

3.6.6. GroEL-Rubisco-ADP·BeF_x and GroEL-Rubisco-ADP·AlF_x (tilt data sets)

UltrAuFoil R2/2 grids were glow discharged at 30 mA for 60 seconds using a Pelco easiGlow system. Three microlitres of GroEL-Rubisco (1 μM) were applied to grids, blotted and plunged into liquid ethane cooled by liquid nitrogen. Plunge-freezing was performed using a Vitrobot mark IV (Thermo Fisher Scientific) operating at 100% humidity and 4 °C. Blot time was set to 5 s, blot force set to -10. Movies (50 frames) were collected using the EPU software on a Titan Krios transmission electron microscope (Thermo Fisher Scientific) operating at 300 keV, equipped with a Gatan K3 direct electron detector operating in super-resolution mode and Gatan energy filter. The defocus range was set between -1.5 and -2.7 μm, and the total exposure was 51 electrons/Å^2. The pixel size was 0.53 Å/pixel in the super-resolution movies. A stage tilt of 30° was set at the start of image acquisition.

3.6.7. GroEL-Rubisco-ADP·BeF_x (two data sets from Chameleon grids)

Grids were prepared using a Chameleon instrument (SPT Labtech) operated by Michele Darrow. GroEL-Rubisco was concentrated to 7 μM, ADP·BeF_x was added as described in **Section 3.2.1**. Movies (50 frames) were collected using the EPU software on a Titan Krios transmission electron microscope (Thermo Fisher Scientific) operating at 300 keV, equipped with a Gatan K3 direct electron detector in counting mode and Gatan energy filter. The defocus range was set between -

1.5 and -2.7 μm, and the total exposure was 50 electrons/Å^2. The pixel size was 0.53 Å/pixel in the super-resolution movies.

3.6.8. GroEL-GroES-Rubisco-ADP·AlF$_x$ (Chameleon data set)

Grids were prepared using a Chameleon instrument (SPT Labtech) operated by Michele Darrow. GroEL-Rubisco was concentrated to 7 μM, GroES and ADP·AlF$_x$ were added as described in **Sections 3.2.1** and **3.2.2**. Movies (50 frames) were collected using the EPU software on a Titan Krios transmission electron microscope (Thermo Fisher Scientific) operating at 300 keV, equipped with a Gatan K3 direct electron detector in counting mode and Gatan energy filter. The defocus range was set between -1.5 and -2.7 μm, and the total exposure was 72 electrons/Å^2. The pixel size was to 0.43 Å/pixel in the super-resolution movies.

3.7. Image processing

The overall image processing strategy, including the programmes used, was similar or identical between data sets. The overall strategy will be described here to avoid repetition. Then, differences in image processing strategies will be detailed in the relevant results chapters and sections. Image processing of each data set is also depicted in "flowchart" style figures in the results chapters.

Micrograph movies were corrected for beam induced motion using Motioncorr2 (Zheng et al., 2017). For movies collected in super-resolution mode with a Gatan K3 camera, micrographs were down sampled by a factor of 2 during motion correction. The CTF parameters of motion-corrected micrographs were estimated using Gctf (Zhang, 2016). Particles were picked using the neural net particle picker included in EMAN version 2.2 (Tang et al., 2007). Particle coordinates (.box files) were then imported into RELION version 3.1 (Zivanov et al., 2018). Particles were extracted from micrographs usually with 2-3x down-sampling, giving typical pixel sizes of 2.0 – 3.6 Å. We generally used down-sampled particles for initial classification, then re-extracted particles at smaller pixel sizes for focussed classification or 3D refinements. Particles were

imported into cryoSPARC v2 and subjected to at least three rounds of reference-free 2D classification. Particles from featureless, noisy, or poorly resolved classes were discarded. Good particles from 2D classification were imported back into Relion using the csparc2star.py Python script. All subsequent image processing steps were performed in Relion v3.1 (unless stated), and without applying any symmetry. For 3D refinements, an initial model of GroEL or GroEL-GroES was generated from a previously published cryoEM reconstruction (EMDB: 3415 and EMDB: 2325) and low-pass filtered to 60 Å.

After 2D classification, particles were refined against a low-pass filtered reference. This yielded a consensus map, usually at a Nyquist-limited resolution. Then, 3D classification of the whole complex (with or without alignments) was used to identify well-resolved particles. Identifying the best particles typically required several rounds of 3D classification with different parameters.

For focussed 3D classifications (to separate substrate-bound from apo particles), a cylindrical volume was created using SPIDER (Frank et al., 1996) and converted to a soft mask using Relion. The mask was aligned to the GroEL consensus map manually using UCSF Chimera (Pettersen et al., 2004) and positioned to encompass a portion of the GroEL cavity where non-native substrate was expected to bind. The mask encompassed minimal density from the hydrophobic helices H and I, and minimal density from the C-terminal tails of GroEL subunits. 3D classification was then performed without alignments. For 3D classification of small volumes, such as non-native substrate, it was necessary to increase the amplitude contrast value in the partial metadata .star files from 0.1 to 0.3. This resulted in 3D classes with strong density for bound non-native substrate. It is not clear why increasing the amplitude contrast gave better results from focussed 3D classification. It might be related to an increase in contrast at low spatial frequencies. This approach is not without precedent and has been mentioned in a recent cryoEM study (Herzik et al., 2019).

3.8. Model building

Models were built into locally sharpened maps generated by Relion. Starting models were based on the crystal structures of GroEL, GroEL-ATP (Rs1), and GroEL-GroES (PDB: 1SS8, PDB: 4AAQ and PDB: 1SVT respectively). Crystal structures were fit into maps using UCSF Chimera (Pettersen et al., 2004) then refined using the ISOLDE module in ChimeraX (Pettersen et al., 2020). Secondary structure geometry restraints were used in all steps of model building. After refinement in ISOLDE, we refined models using real-space refinement in Phenix (Liebschner et al., 2019). For Phenix real-space refinement we generally used the default settings (secondary structure restraints, Ramachandran restraints, rotamer restraints and refinement of B-factors). However, we also used reference model restraints with the starting model from ISOLDE used as the reference. Additional details on model building is given in the relevant sections. Map-model FSC and map quality was assessed using the cryoEM validation tools in Phenix and Molprobity (Williams et al., 2017). Figures of cryoEM maps and atomic models were generated using UCSF Chimera (Pettersen et al., 2004) and ChimeraX (Pettersen et al., 2020).

3.9. Mass spectrometry

Samples for native mass spectrometry were exchanged into 50 mM ammonium acetate (pH 6.8) using 10 kDa cut-off Amicon Ultra centrifugal filtration units (Merck Millipore). Samples were introduced to a first-generation Waters Synapt QToF (Waters Corporation, UK) in nano electrospray gold-coated borosilicate glass capillaries (prepared in-house). Mass calibration was performed using a solution of 30 mg/mL caesium iodide (Fluka). Typical machine parameters used were: capillary 1.4 kV, sampling cone 150 V, extraction cone 4.5 V, backing pressure 7.5 mbar, trap CE 40 eV, transfer CE 10 eV, bias 88 V, source wave velocity 300 ms^{-1}, source wave height 0.2 V, trap wave velocity 300 ms^{-1}, trap wave height 0.2 V. Spectra were analysed using MassLynx v4.1 (Waters Corporation, UK) and Amphitrite (Sivalingam et al., 2013). Spectra were

loaded into Amphitrite using a grain size of 3 and a smoothing value of 2. Peaks were selected and mass values were calculated.

Chapter IV

4. Formation of GroEL-substrate complexes

This chapter describes the preparation and characterisation of the different GroEL complexes we studied by cryoEM. We confirmed the formation of GroEL-Rubisco binary complexes by native PAGE and native mass spectrometry. Native PAGE results were ambiguous and showed multiple species of GroEL-Rubisco (**Section 4.1**). Native mass spectrometry experiments showed one Rubisco monomer bound per GroEL tetradecamer (**Section 4.2**). We also formed stable substrate-bound complexes bound to different nucleotides with or without GroES (**Section 4.3**).

4.1. Formation and characterisation of GroEL-Rubisco complexes

Binary complexes were prepared by diluting denatured Rubisco into a GroEL-containing buffer at a final Rubisco (monomer) to GroEL (tetradecamer) molar ratio of 4:1. After a ten-minute incubation period, aggregated protein was removed by centrifugation. Samples were analysed by blue native polyacrylamide gel electrophoresis (BN-PAGE; **Fig. 4-1**). Apo-GroEL produced a single band near the 720 kDa protein marker (**Fig. 4-1**; lane 1). Since migration of proteins in BN-PAGE is determined by several factors (shape, molecular weight, charge), marker proteins are not true indicators of molecular weight as they are in SDS-PAGE. The single band produced by apo-GroEL therefore corresponded to its 800 kDa tetradecameric structure. Folded Rubisco produced a single band corresponding to its 101 kDa native dimeric form (**Fig. 4-1**; lane 2). GroEL-Rubisco binary complexes gave rise to three bands (**Fig. 4-1**; lane 3). The lowest band was identical to that of apo-GroEL (**Fig. 4-1**; lane 3). The faint middle and uppermost bands represented binary complexes of GroEL bound to non-native Rubisco (**Fig. 4-1**; lane 3, red and blue arrowheads).

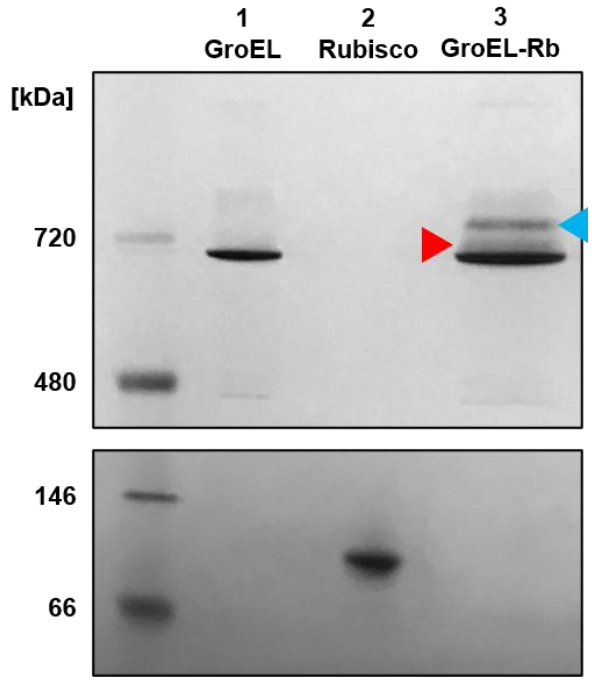

Figure 4-1. Formation of GroEL-Rubisco binary complexes. BN-PAGE of apo GroEL (lane 1), native Rubisco (lane 2) and GroEL-Rubisco binary complexes (lane 3). Red and blue arrowheads in lane 3 mark the position of the two additional GroEL-Rb bands.

4.2. Native mass spectrometry confirmed a 1:1 stoichiometry between GroEL and non-native Rubisco

Some GroEL substrates such as gp23 or MDH have been shown to bind to both GroEL rings simultaneously (Clare et al., 2009 and van Duijn et al., 2007). Previous native mass spectrometry experiments have shown that Rubisco binds to only one GroEL ring. We confirmed this using native mass spectrometry (**Fig. 4-2**). Binding of non-native Rubisco to a GroEL ring is thought to exert negative cooperativity for substrate-binding to the opposite ring (van Duijn et al., 2006). Apo GroEL gave sharp m/z peaks with a calculated mass of 801 kDa (**Fig. 4-2A**). The spectra of GroEL-Rubisco showed broader peaks a shift to higher m/z values than apo GroEL (**Fig. 4-2B**). The calculated mass of GroEL-Rubisco was 846 kDa. This is close to the theoretical combined mass of GroEL-Rubisco (851.5 kDa), i.e., one Rubisco monomer (50.5 kDa) bound to one GroEL tetradecamer (801 kDa). The 5.5 kDa discrepancy was likely due to the poorly resolved peaks for the GroEL-Rubisco complex. This was possibly a result of lower ionisation efficiency or only partial occupancy or weak binding of non-native Rubisco. Native Rubisco produced sharp peaks and had a calculated mass of 101 kDa, consistent with its dimeric form (**Fig. 4-2C**).

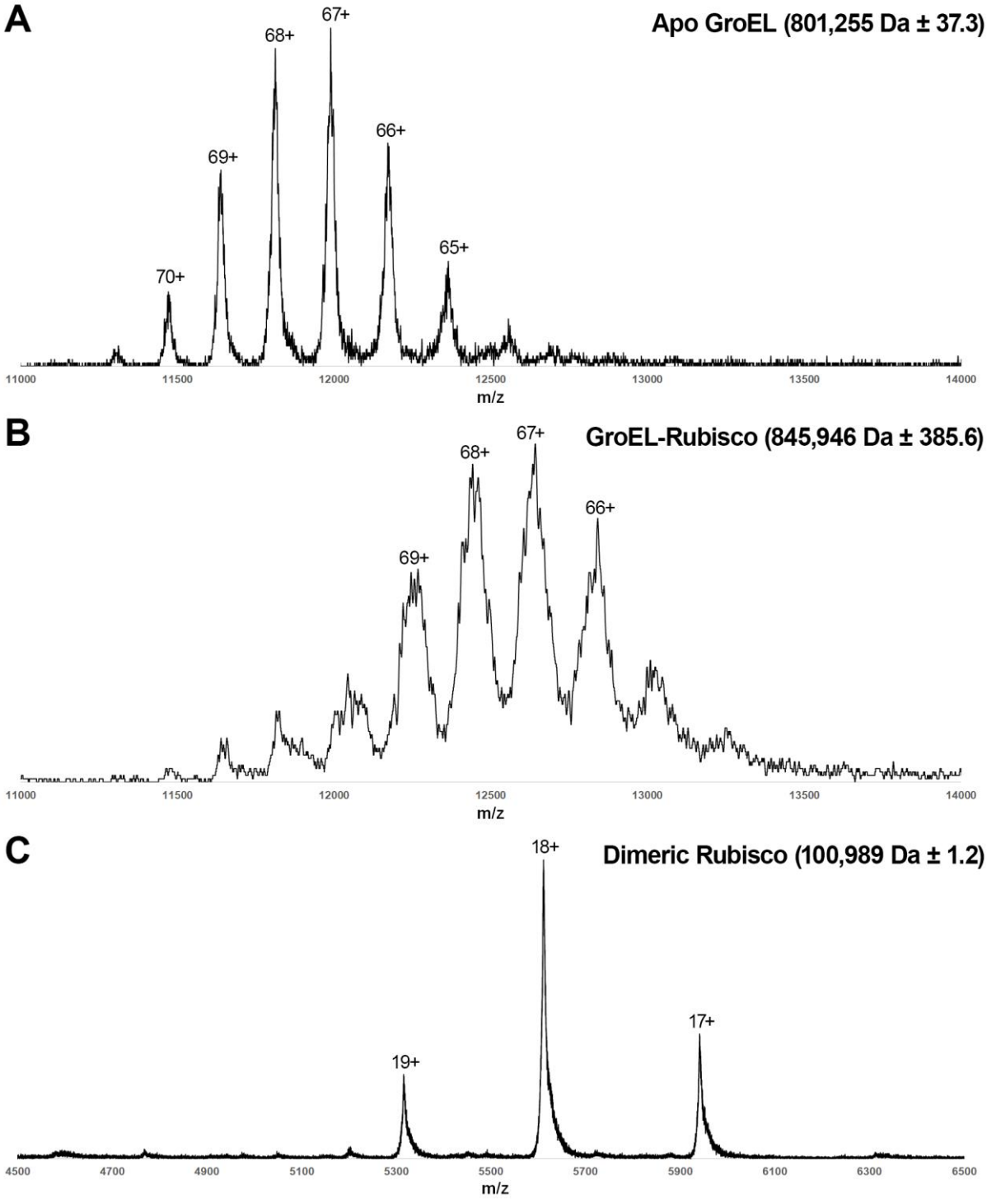

Figure 4-2. ESI mass spectra of apo GroEL, GroEL-Rubisco and native dimeric Rubisco. Nano-ESI of (**A**) Apo GroEL, (**B**) GroEL-Rubisco and (**C**) Native Rubisco. The charge state distribution of each spectrum is indicated and only labelled peaks were used for mass calculations. Masses were calculated using the Amphitrite software and spectra are shown without smoothing.

4.2.1. The discrepancy between native PAGE and native mass spectrometry experiments

Native mass spectrometry showed 1:1 binding between non-native Rubisco and GroEL. However, we observed two gel bands for GroEL-Rubisco in native PAGE experiments. One explanation for this is that two Rubisco monomers bind simultaneously to GroEL, but that one dissociates during native mass spectrometry experiments. A difference in the binding affinities of the two Rubisco monomers, perhaps caused by negative cooperativity, could explain this. The less tightly bound monomer might dissociate during ionisation of the complex. A second explanation is that both gel bands represented the same molecular weight ($GroEL_{14}$-$Rubisco_1$ – 851.5 kDa). There could have been two species of GroEL-Rubisco with different electrophoretic properties. These would be indistinguishable by native mass spectrometry without ion mobility. Rubisco could have bound to GroEL in multiple conformations. These conformations of bound Rubisco may have been different enough to separate by native PAGE. Previous studies have shown that substrates can bind multivalently to either consecutive or non-consecutive GroEL subunits (Farr et al., 2000 and Elad et al., 2007). The two GroEL-Rubisco bands were seen in all repeats of BN-PAGE experiments (see **Fig. 4-3**) and might represent these two different substrate-binding modes. When a substrate binds to consecutive GroEL subunits it is restricted to one side of the cavity, perhaps in a more compact conformation. When a substrate binds to non-consecutive GroEL subunits, it might fill more of the cavity or be suspended across its centre. The more extended conformation of bound non-native Rubisco could have caused slightly slower migration in BN-PAGE.

4.3. Formation of nucleotide and GroES-bound GroEL-Rubisco complexes

Next, we studied the effects of nucleotide and GroES binding on GroEL-Rubisco. We aimed to determine which complexes of GroEL-Rubisco-nucleotide and GroEL-GroES-Rubisco-nucleotide could be stably formed for cryoEM experiments. We added ADP, ATP, or an ATP-analogue to preformed GroEL-Rubisco binary complexes with or without GroES. Complexes were incubated at 21°C for 2 hours then analysed by BN-PAGE (**Fig. 4-3**).

GroEL-Rubisco binary complexes were formed as described previously, using a 4-fold molar excess of Rubisco over GroEL oligomers. Binary complexes were analysed pre- and post-centrifugation and produced the same two additional bands that were seen previously (**Fig. 4-3**; lanes 2 & 3, red and blue arrowheads). The addition of ADP, with or without GroES, had no observable effects on migration of GroEL-Rubisco (**Fig. 4-3**; lanes 4 & 9). Addition of ATP, presumably resulting in a cycling GroEL complex, removed the lower of the two GroEL-Rubisco bands (**Fig. 4-3**; lane 5). This suggested dissociation of a portion of bound non-native Rubisco. The uppermost GroEL-Rubisco band remained even after a 2-hour incubation with ATP (**Fig. 4-3**; lane 5, yellow arrowhead). This band also exhibited marginally slower migration compared to the uppermost band in nucleotide-free GroEL-Rubisco (**Fig. 4-3**; compare migration of bands marked by green and yellow arrowheads; the same effect is seen in lanes 7 and 8, see bands marked by black arrowheads). These results were identical when ATP was replaced with the ground-state analogue ADP·BeF_x or the ATP hydrolysis transition-state analogue, ADP·AlF_x (**Fig. 4-3**; lanes 6 & 7). The combined addition of ATP and GroES generated folding-active complexes. This resulted in complete dissociation of Rubisco from GroEL, leaving only the apo GroEL band (**Fig. 4-3**; lane 10). A faint band corresponding to the native dimeric Rubisco was also observed in this lane only (**Fig. 4-4**). ADP·BeF_x has been reported to form stable GroEL-$GroES_2$ complexes (Fei et al., 2014). However, we did not see this by native PAGE (**Fig. 4-4**; lane 11). Similarly, previous

work has used ADP·AlF_x to form indefinitely stable ternary complexes suitable for X-ray crystallography and cryoEM (Chaudhry et al., 2004 and Clare et al., 2009). In our native PAGE experiments, GroEL-Rubisco incubated with ADP·AlF_x and GroES resulted in a markedly slower migration, suggestive of a significant change in molecular weight (**Fig. 4-4**; lane 12). We attributed this to the formation of stalled GroEL-GroES ternary complexes.

Lane	1	2	3	4	5	6	7	8	9	10	11	12
GroEL	+	+	+	+	+	+	+	+	+	+	+	+
dRubisco	-	+	+	+	+	+	+	+	+	+	+	+
GroES	-	-	-	-	-	-	-	+	+	+	+	+
ADP	-	-	-	+	-	-	-	-	+	-	-	-
ATP	-	-	-	-	+	-	-	-	-	+	-	-
ADP·BeF_x	-	-	-	-	-	+	-	-	-	-	+	-
ADP·AlF_x	-	-	-	-	-	-	+	-	-	-	-	+

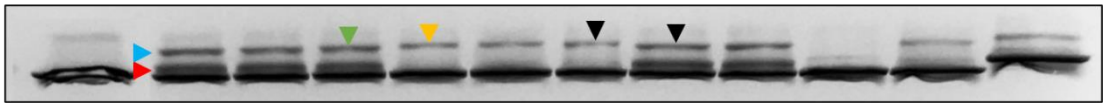

Figure 4-3. Effects of nucleotides and GroES on GroEL-Rb binary complexes. GroEL-Rubisco binary complexes were formed by diluting non-native Rubisco (dRubisco; final concentration 4 μM) into HKM buffer containing 1 μM GroEL oligomers**.** ADP, ATP or ATP analogues were added to a final concentration of 3 mM. GroES was added to a final concentration of 2 μM. Red and blue arrowheads indicate the two species of GroEL-Rubisco. The difference in migration between GroEL-Rubisco and GroEL-Rubisco-ATP is highlighted by the green and yellow arrowheads. Black arrowheads highlight this same difference in migration.

GroEL-Rb	+	+
GroES	+	+
ADP	+	-
ATP	-	+

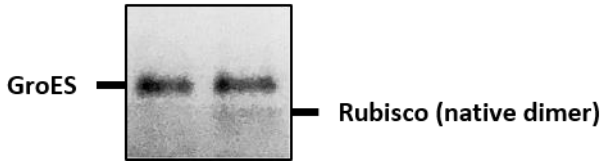

Figure 4-4. Reconstitution of dimeric Rubisco by GroEL/GroES/ATP. GroES and dimeric Rubisco bands from lanes 9 & 10 in **Fig. 4-3**. Note that dimeric Rubisco (101 kDa) consistently migrated faster than native GroES (70 kDa) during BN-PAGE, perhaps due to a more compact conformation.

4.4. Summary and conclusion

We used BN-PAGE to characterise GroEL complexes and confirm their stability for cryoEM experiments. We observed two stable species of GroEL-Rubisco with different electrophoretic properties. In the absence of GroES, one species (the faster migrating band) disappeared following the addition of ATP or an ATP analogue. This species might represent GroEL weakly bound to a non-native Rubisco monomer that dissociates upon ATP binding to GroEL. This may seem counterproductive as it would presumably lead to aggregation. However, *in vivo*, GroES would likely be in close proximity and bind to GroEL before the substrate dissociated and aggregated.

Our BN-PAGE results seem to suggest a 1:1 stoichiometry between non-native Rubisco monomers and GroEL tetradecamers. If two Rubisco monomers had bound to GroEL, presumably doubly-bound GroEL would migrate more slowly than singly-bound GroEL. If one of the Rubisco monomers dissociated from GroEL upon ATP binding, we would only have seen the loss of the slower-migrating band. However, in our experiments it is the faster-migrating species that is lost after ATP binding. The two GroEL-Rubisco bands therefore seem more representative of weak vs strong Rubisco binding, rather than single vs double Rubisco binding.

Nevertheless, our results confirmed the formation of stable GroEL-Rubisco binary complexes. Additionally, substrate-bound complexes of GroEL-ATP or GroEL-ATP-analogue were stable for more than 2 hours at room temperature. Lastly, we formed indefinitely stable GroES-bound complexes with ADP·AlF_x. We therefore decided to use ADP·AlF_x and not ADP·BeF_x for our structural studies of the ternary complex.

Chapter V

5. CryoEM study of GroEL-Rubisco

BN-PAGE and native mass spectrometry experiments in chapter IV showed formation of stable GroEL-Rubisco binary complexes. We aimed to study these by electron microscopy and single-particle analysis. Initial negative stain EM images showed that GroEL-Rubisco complexes adopted a strongly preferred orientation on the EM grid. An initial cryoEM data set produced a reconstruction of GroEL-Rubisco. However, the preferred orientation meant Rubisco was poorly resolved (**Section 5.1**). We attempted to overcome the preferred orientation using a variety of biochemical and grid-based methods. (**Section 5.2**). Unsuccessful in our attempts, we opted to collect cryoEM data set from a tilted sample. This approach degraded resolution but yielded two similar reconstructions of GroEL-Rubisco (**Section 5.3**). We collaborated with SPT Labtech and used their Chameleon instrument to prepare grids for cryoEM. Grids prepared using the Chameleon yielded an improved reconstruction of GroEL-Rubisco. We identified GroEL residues interacting with non-native substrate (**Section 5.4**).

5.1. The presence of non-native Rubisco caused GroEL complexes to adopt a preferred orientation on EM grids

Apo GroEL and GroEL-Rubisco were analysed by negative stain EM. Samples were applied to carbon-coated EM grids and stained with 2% uranyl acetate. Apo GroEL particles mainly adopted end and side orientations (**Fig. 5-1A**). Binary complexes of GroEL-Rubisco exhibited a strongly preferred orientation in which most particles adopted end views (**Fig. 5-1B**). Micrographs of GroEL-Rubisco also showed a higher concentration of particles compared to apo GroEL (**Fig. 5-1B**). Grids of both samples were prepared using identical solution concentrations and grid-preparation methods. The concentrating effect of GroEL-substrate was consistent across multiple grids and sessions. The presence of non-native Rubisco likely caused the complex to become "sticky" and adsorb more strongly to the carbon film on the EM grid. Presumably, non-native substrate bound to GroEL creates a highly hydrophobic end that promotes its adsorption to the carbon surface of the EM grid. Detergents such as octyl glucoside failed to decrease the preferred orientation of GroEL-Rubisco (**Fig. 5-1C**). We had some success when we applied 0.01% poly-L-lysine to the grids prior to sample application. Poly-L-lysine-treated grids reduced the preferred orientation of GroEL-Rubisco and increased the proportion of side views (**Fig. 5-1D**). Contrast was reduced in the resulting images and particles were further concentrated on the support film (**Fig. 5-1D**; inset). Because negative stain EM offers limited information on internal structure, it was not clear if these side views represented intact GroEL-Rubisco complexes. Or, had the poly-L-lysine stripped GroEL of its bound Rubisco? We tried to reproduce these results in cryoEM experiments, without success.

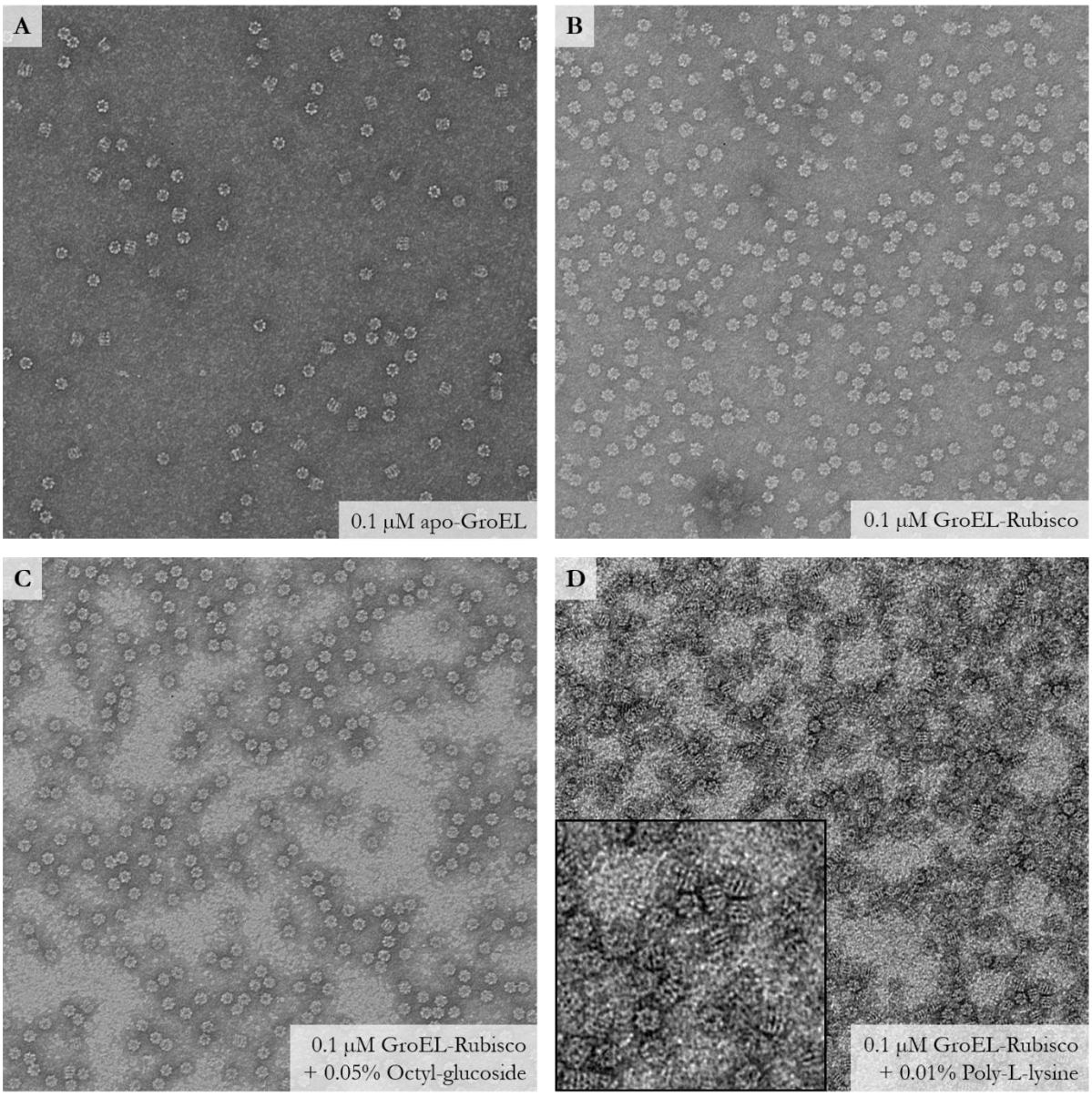

Figure 5-1. Negative stain electron microscopy. Samples of apo GroEL or GroEL-Rubisco were prepared for negative stain EM. Octyl glucoside was added to sample immediately prior to grid preparation. Poly-L-lysine (polyK) grids were prepared by application of 0.01% polyK to grids, allowed to adsorb for 1 minute before blotting away of excess solution. PolyK-coated grids were dried for at least ten minutes before sample application.

5.1.1. The preferred orientation of GroEL-Rubisco in cryoEM

We prepared apo GroEL for cryoEM. An asymmetric reconstruction of apo GroEL from 42k particles yielded a 4.3 Å map (**Fig. 5-2A**). Apo GroEL showed features consistent with previously

published crystal structures and cryoEM reconstructions. Apo GroEL particles adopted a range of views, but most represented side views of GroEL (**Fig. 5-2B**).

Substrate-bound GroEL was prepared for cryoEM by dilution of urea-unfolded Rubisco into a GroEL-containing buffer. A 3.4 Å map was reconstructed from 300k particles (**Fig. 5-2C**, grey volume). No substrate density was visible in the initial consensus map. Presumably, there was both compositional and conformational heterogeneity present, as seen in our native page experiments in Chapter IV. Compositional heterogeneity arose from the presence of both apo GroEL and GroEL-Rubisco in the same sample. Additionally, non-native substrates adopt an ensemble of conformations with varying degrees of structure. Heterogeneity in cryoEM leads to less well resolved features that become averaged or "blurred" out in a consensus refinement. 3D classification is used to identify subsets of particles with more well-resolved local features. Default 3D classification of GroEL-substrate complexes in Relion failed to separate apo from substrate-bound GroEL. The mass of GroEL (800 kDa) accounts for 94% of a GroEL-Rubisco complex. The signal originating from GroEL dominates both the alignment and classification steps of 3D classification. We therefore used a focussed 3D classification strategy targeted at the GroEL cavity. A cylindrical mask that encompassed the GroEL cavity was created and used in 3D classification. (**Fig. 5-2C**, blue volume). Using 3D classification in Relion, particles were assigned to one of three classes (**Fig. 5-2D**). Class 1 showed stronger central density than the other two classes, and this density presumably represented non-native Rubisco (**Fig. 5-2D**, left). We refined particles from class 1 and calculated a reconstruction of GroEL-Rubisco (**Fig. 5-2E**). The angular distribution of GroEL-Rubisco particles was narrower than apo GroEL and most particles adopted end views (**Fig. 5-2F**). The map showed only weak density for bound substrate. We fitted the crystal structure of a previously published model of apo GroEL (PDB: 1SS8). Non-native Rubisco appeared to be bound between three consecutive GroEL subunits, and two opposite consecutive GroEL subunits (**Fig. 5-2G**). This conformation of bound non-native Rubisco was somewhat similar to a previously published cryoEM reconstruction of GroEL-Rubisco (Natesh et al., 2018). Contacts

between non-native Rubisco and GroEL were located in both hydrophobic helices of the GroEL apical domains (**Fig. 5-2H**). However, we deemed the resolution insufficient to identify exact contact residues. The density for non-native Rubisco accounted for ~33% the volume of a folded Rubisco monomer.

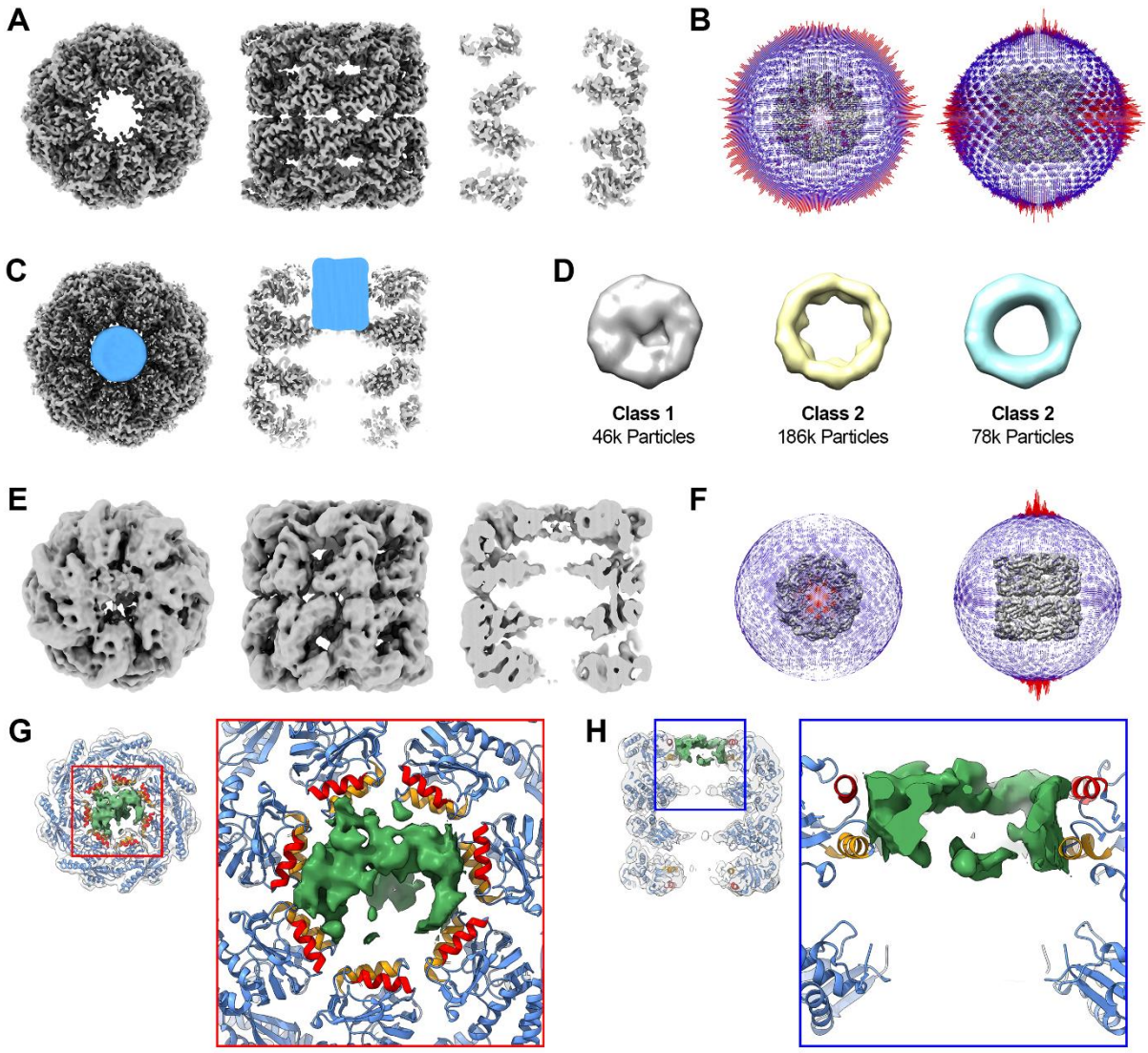

Figure 5-2. Initial cryoEM data sets and image processing of apo GroEL and GroEL-Rubisco. (**A**) Reconstruction of apo GroEL from 42,215 particles. Map shown has been sharpened with a negative B-factor. (**B**) 3D Histogram showing the angular distribution for apo GroEL particles. Each cylinder represents a discrete Euler angle and its length and colour (scaled from blue to red) corresponds to the number of particles assigned to that angle. Angular

distribution plots were generated automatically during 3D refinement in RELION and viewed in UCSF Chimera (Pettersen et al., 2004). (**C**) Focussed classification of the GroEL-Rubisco cavity. A 3D mask (blue) was created and positioned to encompass the presumed substrate-binding region of the GroEL cavity. (**D**) 3D classification (without alignments) of the masked region was performed in Relion v2.1. 3D Classes representing apo GroEL (class 2 and class 3) and substrate-bound GroEL (class 1) particles were calculated. (**E**) CryoEM reconstruction of class 1 particles. (**F**) 3D Histogram showing the angular distribution for GroEL-Rubisco particles. (**G**) Close-up end view of non-native Rubisco bound to GroEL. Density representing GroEL was erased and replaced with the fitted crystal structure of apo GroEL (PDB: 1SS8). (**H**) Closeup side view section of non-native Rubisco bound to GroEL.

5.1.2. Testing the extent of preferred orientation of GroEL-Rubisco particles in cryoEM

We were concerned that the side and oblique views in the GroEL-Rubisco cryoEM data set were in fact mis-classified particles of apo GroEL. To test this, we calculated a reconstruction using only the side view particles from class 1. We used 2D classification in Relion to identify and group side view particles from class 1. Following 3D refinement, the reconstruction from side view particles showed no significant substrate density, even at low threshold values (**Fig. 5-3A**). The orientation distribution plot confirmed that only side view particles contributed to the reconstruction (**Fig. 5-3B**). This supported the idea that there were two populations of GroEL particles in the sample. Apo GroEL particles adopted a range of views, including end and side views. Substrate-bound GroEL particles had a near-absolute preference for adopting end views.

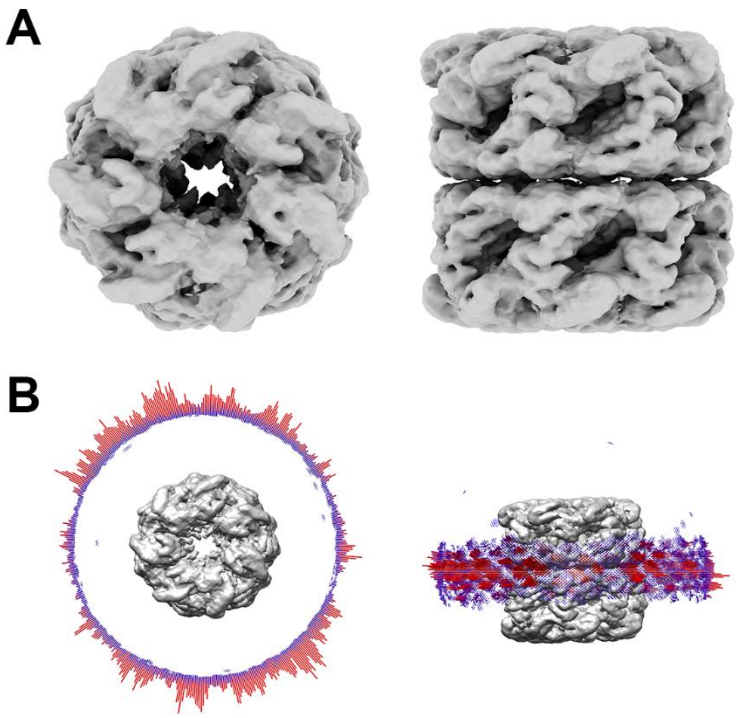

Figure 5-3. Reconstruction of side view particles from 3D class 1. (**A**) Particles from 3D class 1 were subjected to reference-free 2D classification in Relion. Side view classes were picked and particles from these classes were grouped. The new group of side view particles was refined. (**B**) 3D histogram showing the angular distribution for side-view particles.

5.1.3. Consequences of the preferred orientation of GroEL-Rubisco

Substrate-bound GroEL complexes have been known to preferentially adopt end views in cryoEM (Elad et al., 2007). We had hoped that the latest improvements in cryoEM hardware and software would enable us to produce an improved reconstruction of a GroEL-substrate complex. However, this was not the case. Poorly resolved density for bound non-native Rubisco was partly due to the preferred orientation of GroEL-Rubisco. Our reconstruction of GroEL-Rubisco in **Fig. 5-2E** represented a combination of apo GroEL and GroEL-Rubisco. We initially tried to further classify particles from class 1 using the same mask as in **Fig. 5-2C**. However, the new classes of particles yielded poorly resolved reconstructions following 3D refinement (not shown). The preferred

orientation GroEL-Rubisco had to be addressed as it made a high-resolution 3D reconstruction of the complex impossible. Previous experience told us that we would encounter similar problems when studying the nucleotide bound GroEL complexes.

5.2. Strategies to overcome preferred orientation in EM

Preferred orientation is a common problem in cryoEM. It arises when proteins adsorb to the air-water interface or to the support film of the grid. Many different strategies have been employed to mitigate it (Drulyte et al., 2018). For example, detergents can be added to the sample prior to grid preparation. Detergents alter the surface tension of the buffer and can in some cases decrease particle adsorption to the air-water interface (e.g. as in Chen et al., 2019). Another approach is to add a support layer to the EM grid. These can affect the orientation distribution of particles by providing another surface onto which particles can adsorb. The chemistry of these surfaces can also be altered to produce different orientation distributions (Naydenova et al., 2019). Amorphous carbon is a commonly used support for EM grids, and in recent years graphene derivatives have also seen widespread use. These approaches are discussed in greater detail in Chapter II (see **Section 2.3.6**).

In practice, biochemical and grid-based approaches to alter the orientation distribution of particles are sample dependent. What works for one sample may fail for another. We experimented with many different approaches to try to solve the preferred orientation of GroEL-Rubisco. None were successful (**Table 5-1** summarises our efforts).

Approach	**Experimental details**	**Effect on orientation distribution**
Additive	300, 600 mM Urea	No effect.
Additive	0.1%, 0.01% Poly-L-lysine	Side views more present in negative stain EM. Effect not reproducible in cryoEM.
Additive + Support layer	0.01% Poly-L-lysine added to grids coated with a monolayer of graphene oxide.	No effect.
Detergent	0.005%, 0.05%, 0.1% v/v Tween-20	No effect.
Detergent	0.005%, 0.05%, 0.1% v/v Triton-X	No effect.
Detergent	2 mM CHAPSO 4 mM CHAPSO 8 mM CHAPSO	No effect.
Detergent	0.05%, 0.1% v/v Octyl glucoside	No effect.

Grid treatment	Positive glow discharge (glow discharge in the presence of amylamine).	No effect.
Grid type	Experimented with different grid hole sizes and hole spacings. C-flat grids, Quantifoil grids, UltrAufoil grids.	No effect.
Support layer	Graphene oxide monolayer	No effect.
Vitrification parameters	Varied blot time, blot force, drain time, wait time, temperature, and humidity.	No effect.

Table 5-1. Attempts to improve the preferred orientation of substrate-bound GroEL. Binary complexes of GroEL-Rubisco were prepared as described in the methods. Various approaches were tested to improve the preferred orientation of substrate-bound GroEL particles in negative stain EM and/or cryoEM. Micrographs were collected on 100 kV or 120 kV electron microscopes with a 4k Gatan CCD camera. Particles were picked in EMAN2, extracted, and analysed by 2D classification either using EMAN2 or Relion. The proportion of end and side view classes was assessed (we were not usually able to identify oblique views in negative stain EM or low resolution cryoEM). If the proportion of side views was < 10% the approach was considered ineffectual. Grids for negative stain EM and cryoEM were prepared in duplicate or triplicate. Where mentioned, the monolayer of graphene oxide was applied to grids using a published method (Cheng et al., 2020).

5.3. Addressing the preferred orientation of GroEL-Rubisco through tilted data collection

The simplest sample-independent method to reduce preferred orientation is to collect cryoEM images at a high tilt angle. In **Section 5.2**, data was collected at 0° tilt (i.e. untilted data). Data collected at higher tilt angles (30 – 50°) limits high resolution information but can improve map quality (Tan et al., 2017).

Binary complexes were prepared for cryoEM on grids coated with a monolayer of graphene oxide. The use of graphene oxide allowed us to use a lower GroEL solution concentration (~50 nM) for

grid preparation. A nominal stage tilt angle of 30 or 40° was applied prior to data collection. Approximately 3,500 micrograph movies were collected at 30° tilt and 1,500 micrographs at 40° tilt. Due to the presence of the graphene oxide monolayer and the use of stage tilt, contrast between GroEL particles and the vitreous ice background was low (**Fig. 5-4A**, top). Particles were picked using the neural net particle picker in EMAN v2.2. The co-ordinates of picked particles were imported into Relion and used to extract particles from micrographs. Extracted particles were imported into cryoSPARC. Reference-free 2D classification showed a range of views. Most 2D class averages represented oblique views of GroEL (**Fig. 5-4A**, bottom). Refinement of 157k particles from well-resolved 2D classes yielded a 5.1 Å consensus map, which was improved to 4.2 Å following CTF refinement and further 3D refinement. Refinement was Nyquist-limited at a sampling frequency of 2.1 Å/px. Re-extraction at a higher sampling frequency (1.4 Å/px) yielded a consensus map with a resolution of 4.0 Å (**Fig. 5-4D**). The Euler angle plot showed that the preferred orientation of GroEL-Rubisco particles had been mitigated. Oblique and side views of GroEL were well represented in the consensus data (**Fig. 5-4C)**.

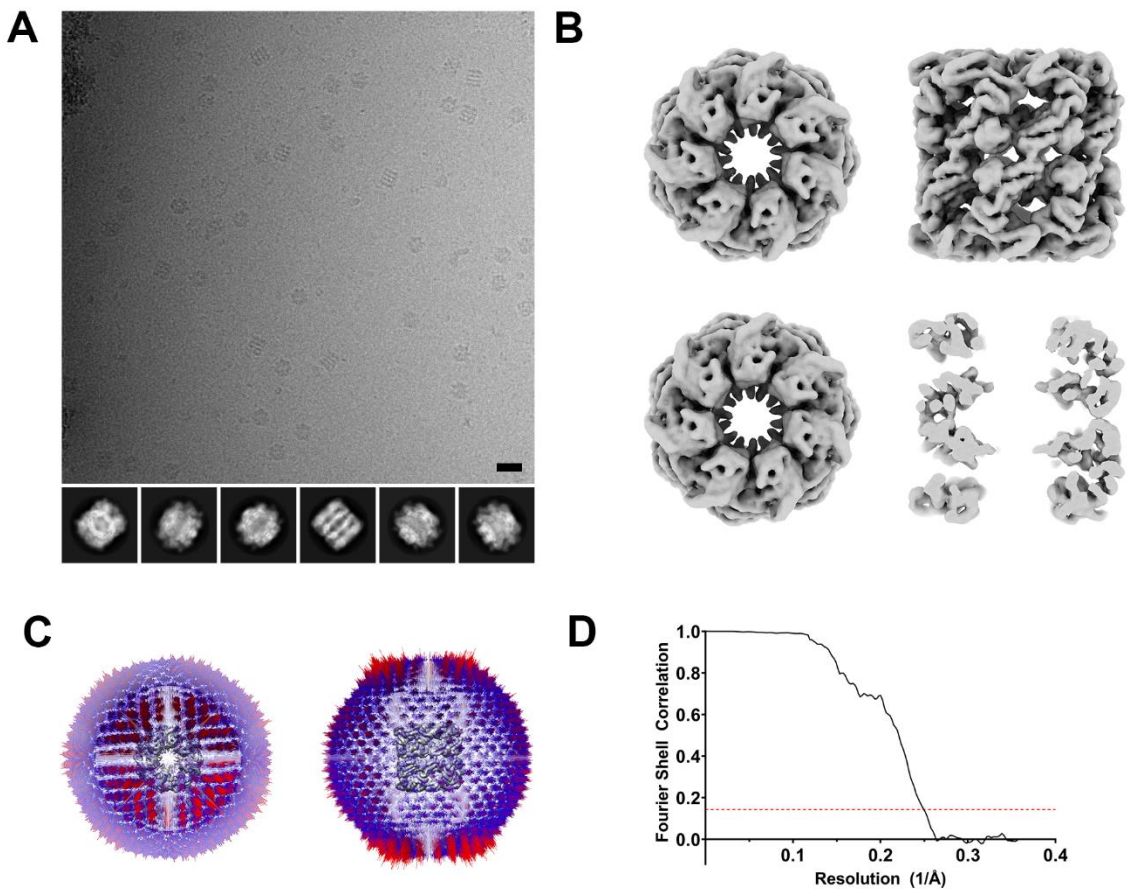

Figure 5-4. CryoEM of GroEL-Rubisco collected at 30 – 40° tilt. (**A**) Representative motion-corrected cryo micrograph (top; black scale bar = 20 nm) and 2D class averages showing a range of views of the GroEL-Rubisco complex (bottom). (**B**) Consensus reconstruction from 157k particles, viewed from the *cis* ring (top left), side (top right), *trans* ring (bottom left) and a central slice through a side view (bottom right). (**C**) 3D histogram showing the orientation distribution of GroEL-Rubisco particles, with the consensus reconstruction shown in the middle of the histogram to illustrate orientation angles. (**D**) Fourier shell correlation (FSC) plot for the calculated half-maps; overall resolution determined by the 0.143 criterion (dashed red line) was 4.0 Å.

5.3.1. Image processing of GroEL-Rubisco cryoEM tilt data set

When viewed at a low threshold level, the 5.1 Å consensus map showed additional density in the GroEL cavities (**Fig. 5-5A**, red arrow). CTF refinement and further 3D refinement improved the resolution of the consensus map to 4.2 Å (**Fig. 5-5B**). Particles were subjected to focussed 3D

classification using a mask that encompassed the GroEL cavity. The approach was the same one used for classification of the untilted data set. It is also described in greater detail in the methods and was essential for identifying subsets of particles with strong density for bound substrate (see **Section 3.7**). An initial round of focussed 3D classification followed by 3D refinement of each class yielded seven classes of GroEL-Rubisco and one class of apo GroEL (**Fig. 5-5C** and **Fig. 5-5D**). Apo GroEL particles were refined to 4.3 Å (**Fig. 5-5D**, class 3). Classes showing substrate density were refined to 4.6 – 8.2 Å (**Fig. 5-5D**, classes 1, 2, 4 – 8). Class 4 contained the highest number of GroEL-Rubisco particles and displayed strong density for bound non-native Rubisco. We performed further 3D classification on particles in class 4 using the same *cis* cavity mask (**Fig. 5-5E**). We then refined particles from each of these three classes (**Fig. 5-5F**). Refinements of classes 4.1 and 4.2 both showed strong density for non-native Rubisco (**Fig. 5-5F**). Class 4.3 showed relatively weak density for non-native Rubisco, despite containing a majority of the particles (**Fig. 5-5F**, cyan). Further classification of class 4.3 did not yield improvements (not shown). We also tried combining particles from classes 1 and 2 into a single class and refinement them together. Although the reconstruction of the combined class showed strong density for bound substrate, the contacts to GroEL were comparatively less well resolved (not shown).

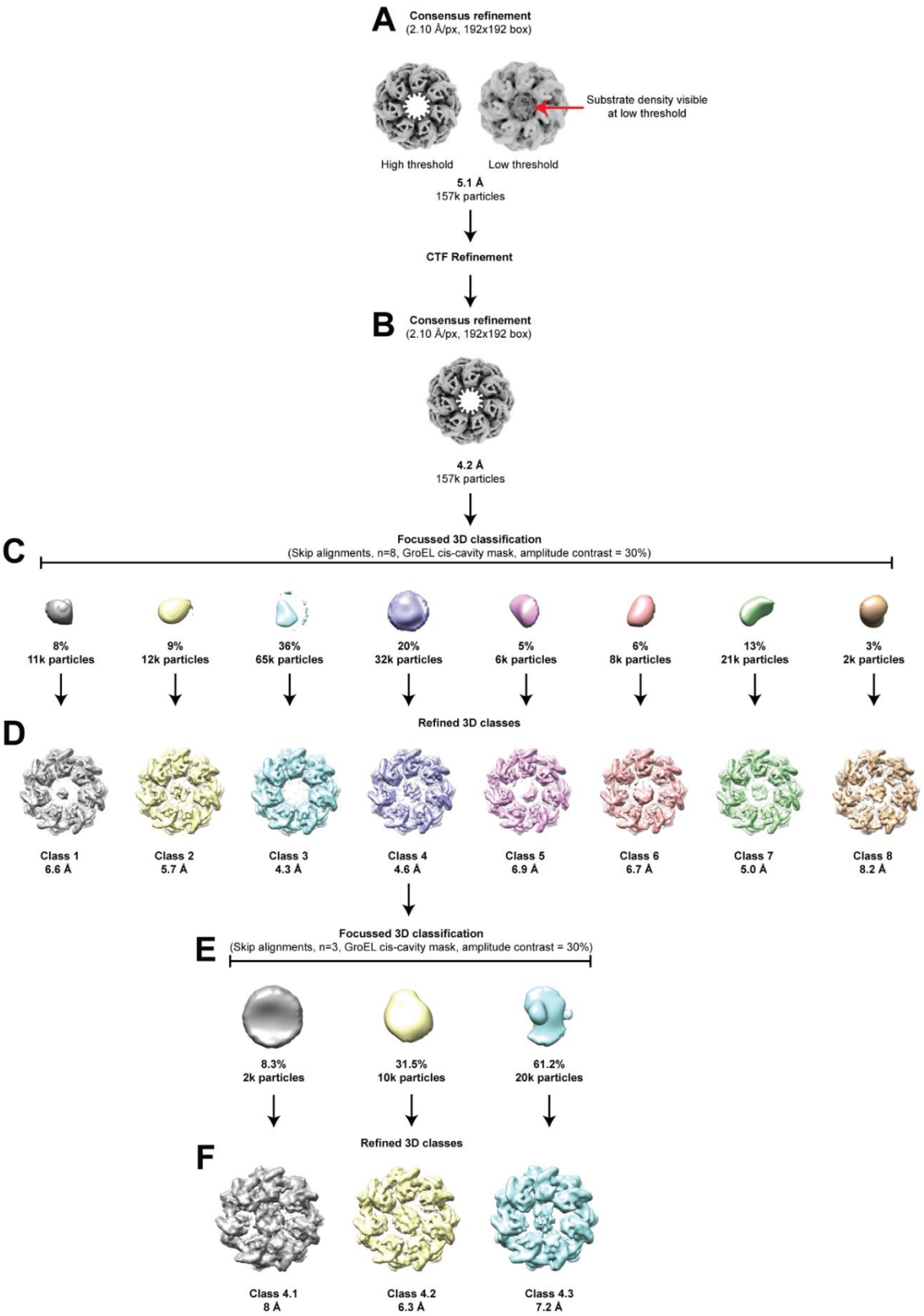

Figure 5-5. Image processing of GroEL-Rubisco tilt data set. (**A**) Consensus refinement from 157k particles. On the right, the same map is shown at a lower density threshold to better illustrate weak substrate density. (**B**) Consensus refinement from 157k following CTF refinement. (**C**)

Focussed 3D classification of substrate density. 3D classification was performed in Relion without alignments. A mask encompassing the *cis* cavity of GroEL was used. Amplitude contrast was set to 0.3 in the input particle metadata .star file. The resulting 3D classes are shown. (**D**) 3D Refinements of particle subsets identified by focussed 3D classification. (**E**) Focussed 3D classification of refined class 4 particles from D. Using the same parameters and mask as in C. (**F**) 3D Refinement of particle subsets identified in E.

5.3.2. Reconstructions of GroEL-Rubisco

Image processing yielded two cryoEM reconstructions of GroEL-Rubisco (**Fig. 5-6**). Both classes showed a well resolved GroEL complex at a moderate resolution and were consistent with previously published structures (**Fig. 5-6A** and **Fig. 5-6B**). Density for bound substrate was clearly visible in the cavity of one GroEL ring (**Fig. 5-6A** and **Fig. 5-6B**, top left). Most GroEL secondary structure was visible in both maps despite the low resolution (**Fig. 5-6A** and **Fig. 5-6B**, top right). No substrate was observed in the opposite GroEL ring (**Fig. 5-6A**, and **Fig. 5-6B**, bottom left). We observed similar topologies for bound substrate in both reconstructions. Rubisco interacted with two consecutive GroEL subunits on one side of the ring and two opposite consecutive GroEL subunits (**Fig. 5-6A** and **Fig. 5-6B**, top left and bottom right). At very low threshold values, weaker interactions with additional GroEL subunits were visible (not shown). Angular distribution plots for both classes suggested the presence of a preferred side view, as well as many absent views (**Fig. 5-6C** and **Fig. 5-6D**). The resolution of classes 1 and 2 was 8.0 and 6.3 Å respectively (**Fig. 5-6E** and **Fig. 5-6F**).

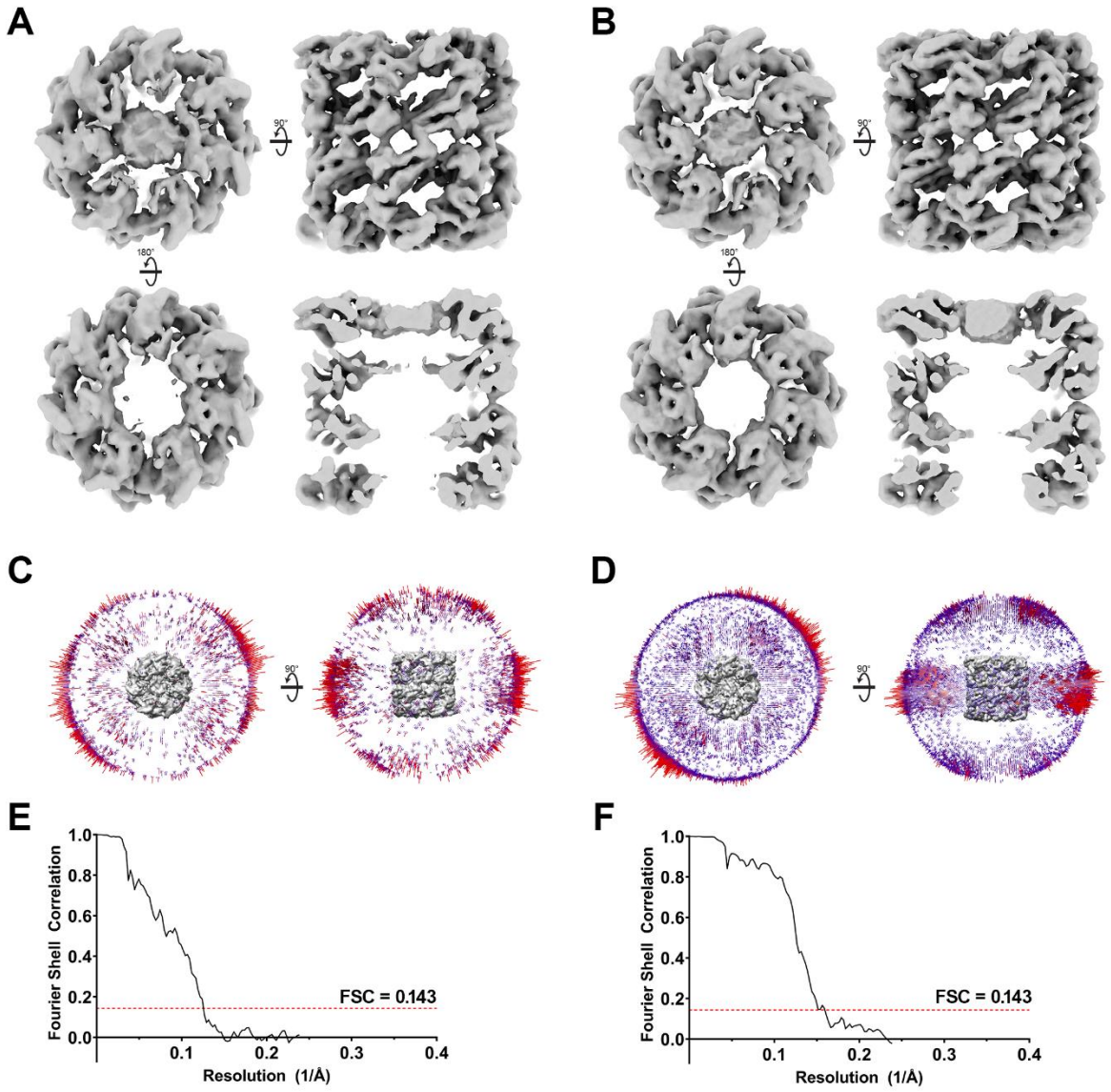

Figure 5-6. Reconstructions of GroEL-Rubisco. 3D refinement of class 4.1 (**A**) and class 4.2 (**B**). 3D histogram showing the orientation distribution of class 4.1 (**C**) and class 4.2 (**D**) GroEL-Rubisco particles, with the class reconstruction shown in the middle of the histogram to illustrate orientation angles. Fourier shell correlation (FSC) plots for the calculated half-maps of class 4.1 (E) and class 4.2 (F). Overall resolutions determined by the 0.143 criterion (dashed red line) were 8.0 Å and 6.3 Å for classes 4.1 and 4.2 respectively.

5.3.3. Contacts between non-native Rubisco and GroEL

To assign contacts between GroEL and substrate, we rigid body fitted a previously published crystal structure of apo GroEL (PDB: 1SS8) to our cryoEM map. The model was optimised in ISOLDE while imposing distance restraints on the individual domains of GroEL. Without distance restraints the apical domain models drifted into the unmodelled substrate density. Although limited by resolution, several GroEL-Rubisco contacts were resolved. In class 1 the main site of contact for non-native substrate was helix I. The residues Arg-268, Val-264 and Thr-261 appeared to show the strongest contacts to the substrate (**Fig. 5-7A** and **Fig. 5-7B**). In one GroEL subunit, Glu257 also appeared to form a contact with Rubisco (**Fig. 5-7A** and **Fig. 5-7B**). The mode of binding and interactions seen in class 2 were similar and appeared to involve the same set of residues (**Fig. 5-7C** and **Fig. 5-7D**). At lower threshold values it became impossible to assign individual contacts with certainty. Substrate appeared to interact weakly with most cavity-facing residues of helix I, and several hydrophobic residues of helix H. Interactions with additional GroEL subunits were also observed at low threshold values.

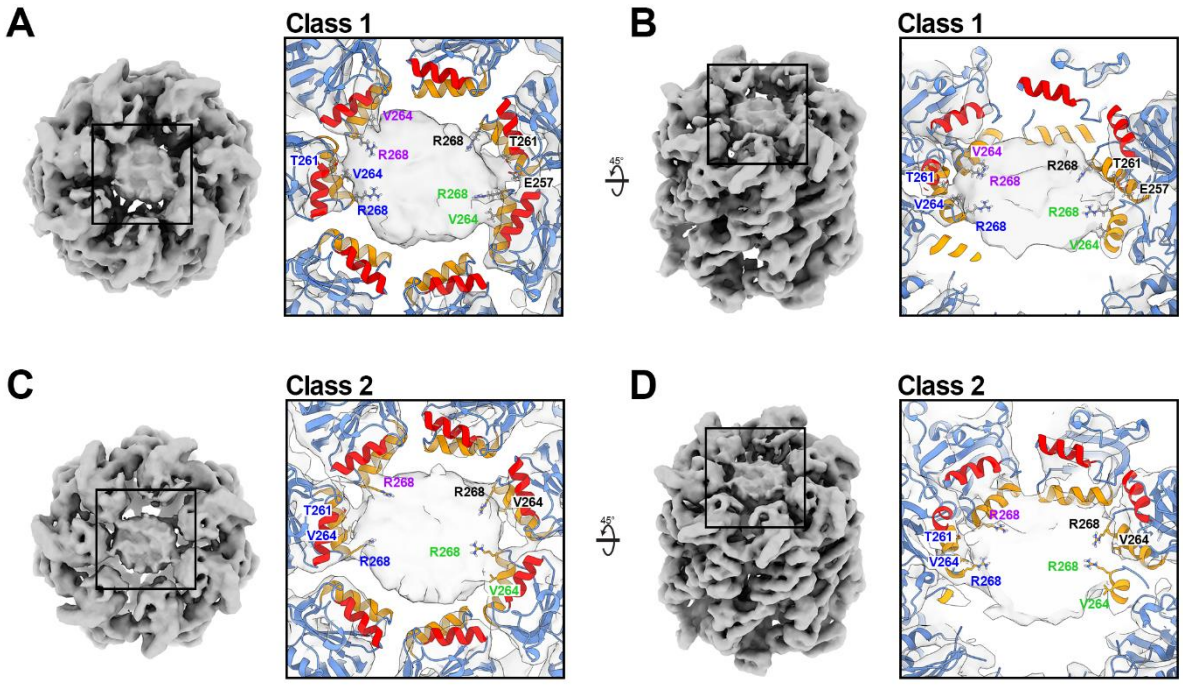

Figure 5-7. Contacts between non-native Rubisco and GroEL apical domains. End and oblique views of the substrate-occupied *cis* ring of GroEL-Rubisco reconstructions from (**A** and **B**) class 4.1 and (**C** and **D**) class 4.2. Interacting residues are indicated and represented by sticks. Residue names that share a colour belong to the same GroEL subunit. GroEL is shown as a ribbon diagram and is coloured blue. Helices H and I are coloured red and orange.

5.3.4. Movement and interactions between adjacent GroEL domains

In class 1, two substrate binding GroEL apical domains appeared to contact each other (**Fig. 5-8**). Despite low resolution, there was continuous density between helix H of both subunits. This interaction could be modelled as the salt bridge Glu232-Cδ-Lys242-Nζ (2.8 Å inter-atomic distance). These salt bridges are not present in apo GroEL. The sidechains involved are > 10 Å apart in the apo GroEL crystal structure (PDB: 1SS8). A similar break in the 7-fold symmetry of GroEL was seen in response to MDH binding in a previous cryoEM study (Elad et al, 2007).

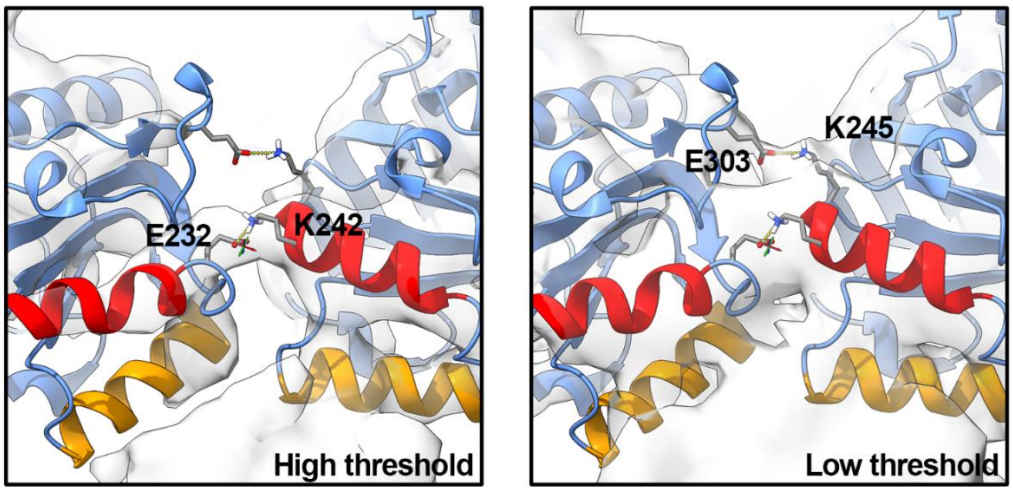

Figure 5-8. Interaction between adjacent GroEL subunits caused by bound non-native substrate. View of adjacent GroEL apical domains with cryoEM density coloured light grey and outlined. CryoEM density is shown at two different thresholds, high (left) and low (right). Residues

involved in the putative salt bridge are shown as sticks and labelled. Helices H and I are coloured red and orange.

5.3.5. Inter-ring contacts

Compared to apo GroEL and to our initial reconstruction of GroEL-Rubisco, the tilt data set reconstructions showed extra density between the two GroEL rings (**Fig. 5-9B**). The extra density was located between the Arg-340 and Glu-434 residues of GroEL subunits in both rings (**Fig. 5-9B**).

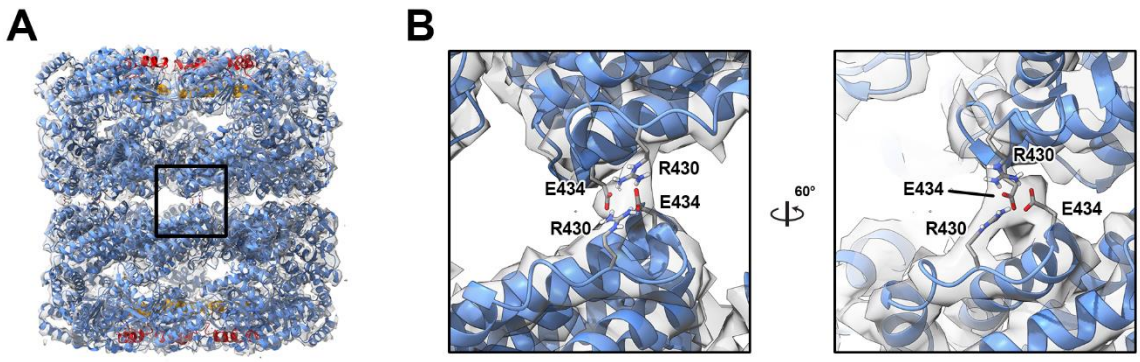

Figure 5-9. Strong inter-ring contacts between substrate-occupied (top ring) and empty (bottom ring) GroEL rings. (**A**) Side view of GroEL-Rubisco class 4.2. The cryoEM density is shown as a transparent light grey. (**B**) Two close views of the inter-ring interaction, highlighting GroEL residues R430 and E434. GroEL is depicted as a ribbon diagram and coloured blue. Helices H and I are coloured red and orange.

5.4. Chameleon data set

In tandem with our efforts to solve the preferred orientation of GroEL-Rubisco through tilt data collection, we explored the use of Chameleon. Chameleon is a next-generation grid preparation instrument for cryoEM (Klebl et al., 2020). The Chameleon instrument itself is discussed in greater detail in the methods (see **Section 3.5**).

5.4.1. Consensus refinements

We prepared grids for cryoEM using the Chameleon at SPT Labtech with the assistance of Michele Darrow. GroEL-Rubisco was prepared as before and concentrated to ~2 mg/mL. Our first Chameleon session yielded grids with dispense-to-freeze times ranging from 50 – 1000 ms. We screened each grid and collected cryoEM data sets from a 101 ms grid (grid 1) and a 1000 ms grid (grid 2). Micrograph movies were motion corrected using MotionCorr2. CTF values were estimation using Gctf. Particles were picked using the neural network particle picker in EMAN v2.2. Particles were extracted in Relion v3.1 then imported into cryoSPARC and subjected to three rounds of 2D classification. Classified particles were imported back into Relion and refined.

Both grids were prepared during the same Chameleon session using the same aliquot of GroEL-Rubisco. However, we initially processed the data from each grid separately (**Fig. 5-10** and **Fig. 5-11**). Few particles were visible on micrographs from either data set (**Fig. 5-10A** and **Fig. 5-11A**, top). Reference-free 2D classification showed a range of views (**Fig. 5-10A** and **Fig. 5-11A**, bottom). Images from grid 1 contained an average of ~70% more particles than grid 2 images (30 and 18 particles per image respectively). Consensus refinements from both data sets yielded similar maps that exhibited density for bound non-native Rubisco (**Fig. 5-10B** and **Fig. 5-11B**). The density for the GroEL complex was quite poor considering the number of particles. For example, in cryoEM experiments using apo GroEL, 30,000 particles are enough to reach a resolution of 4 Å. Side chain density is apparent at this resolution. The orientation distribution suggested that the preferred orientation of GroEL-Rubisco was present in both data sets (**Fig. 5-10C** and **Fig. 5-11C**).

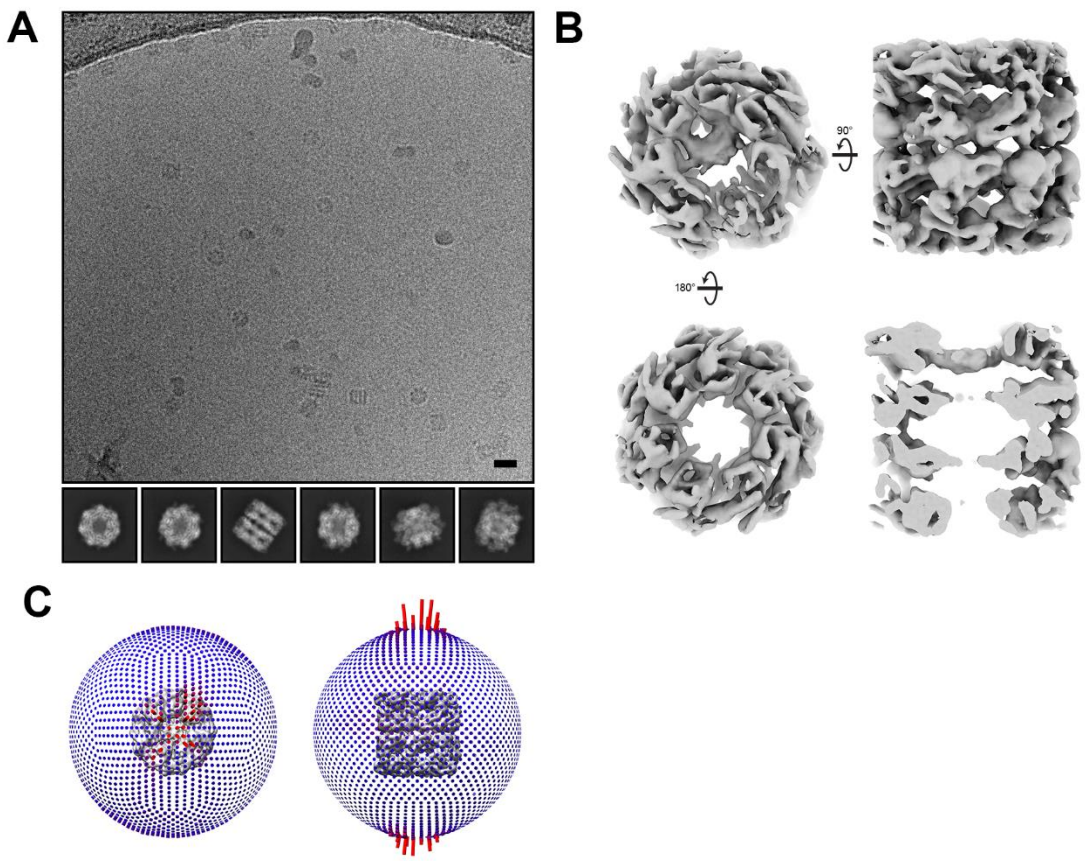

Figure 5-10. Consensus reconstructions from Chameleon grid 1. (**A**) Representative motion-corrected cryo micrograph (top; black scale bar = 20 nm) and 2D class averages showing a range of views of the GroEL-Rubisco complex (bottom). (**B**) Consensus reconstruction from 68k particles, viewed from the *cis* ring (top left), side (top right), *trans* ring (bottom left) and a central slice through a side view (bottom right). (**C**) 3D histogram showing the orientation distribution of GroEL-Rubisco particles, with the consensus reconstruction shown in the middle of the histogram to illustrate orientation angles.

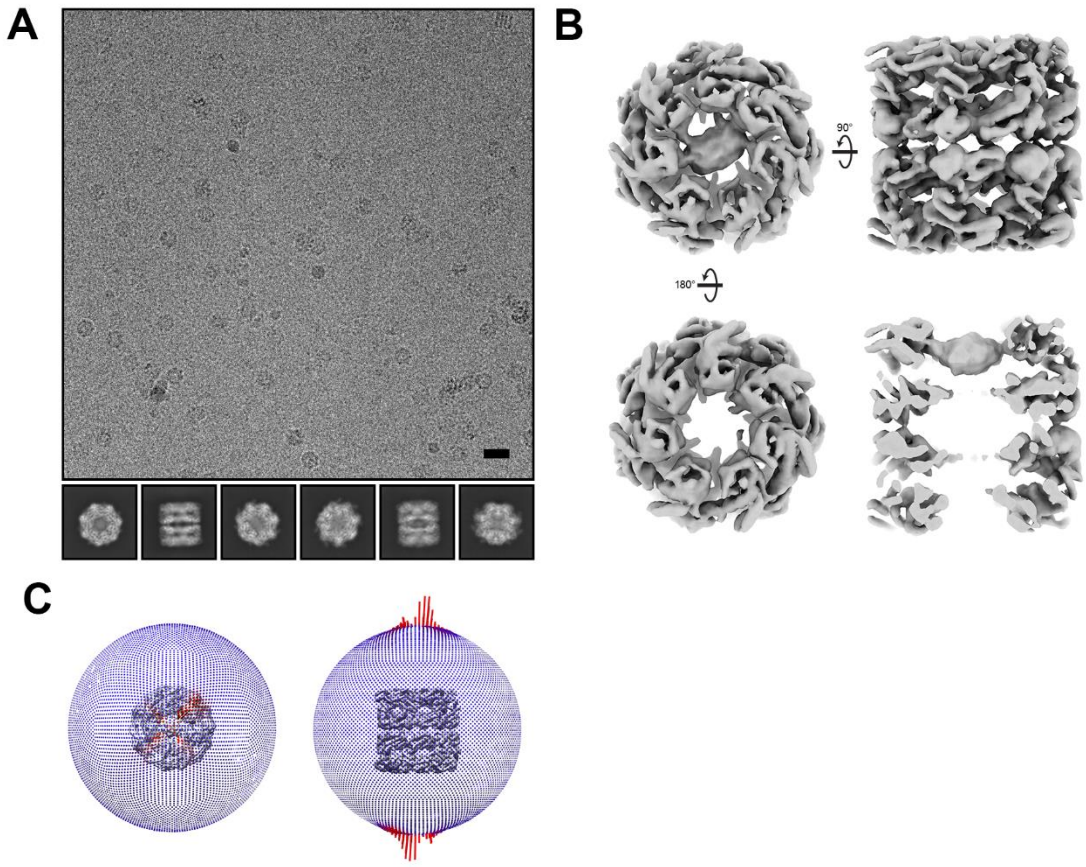

Figure 5-11. Consensus reconstructions from Chameleon grid 2. (**A**) Representative motion-corrected cryo micrograph (top; black scale bar = 20 nm) and 2D class averages showing a range of views of the GroEL-Rubisco complex (bottom). (**B**) Consensus reconstruction from 137k particles, viewed from the *cis* ring (top left), side (top right), *trans* ring (bottom left) and a central slice through a side view (bottom right). (**C**) 3D histogram showing the orientation distribution of GroEL-Rubisco particles, with the consensus reconstruction shown in the middle of the histogram to illustrate orientation angles.

5.4.2. Classification strategy

Refined particles were subjected to 3D classification (**Fig. 5-12**). Interestingly, for these data sets bound substrate was clearly seen in both consensus refinements (**Fig. 5-12A** and **Fig. 5-12B**. Typically, for our cryoEM data sets of GroEL-Rubisco complexes, substrate was only resolved after focussed 3D classification using a mask encompassing the GroEL *cis*-cavity. (This was true

for the nucleotide-bound and GroES-bound complexes presented later in this work, as well as other data sets not presented in this thesis). For this data set of GroEL-Rubisco, the usual focussed 3D classification strategy was not required.

For grids 1 and 2, 3D classification yielded substrate-bound classes of 14k and 51k particles (**Fig. 5-12C** and **Fig. 5-12D**). These classes showed stronger density for bound non-native Rubisco than the consensus map. GroEL-Rubisco particles were merged into a single class, re-extracted in a larger box, and refined (**Fig. 5-12E**). Further classification from the combined class yielded low resolution structures exhibiting anisotropy and poorly resolved features (not shown). Additionally, ~30% of the particles from each data set were identified as apo GroEL. These classes were merged and refined to 4.2 Å (**Fig. 5-12F**).

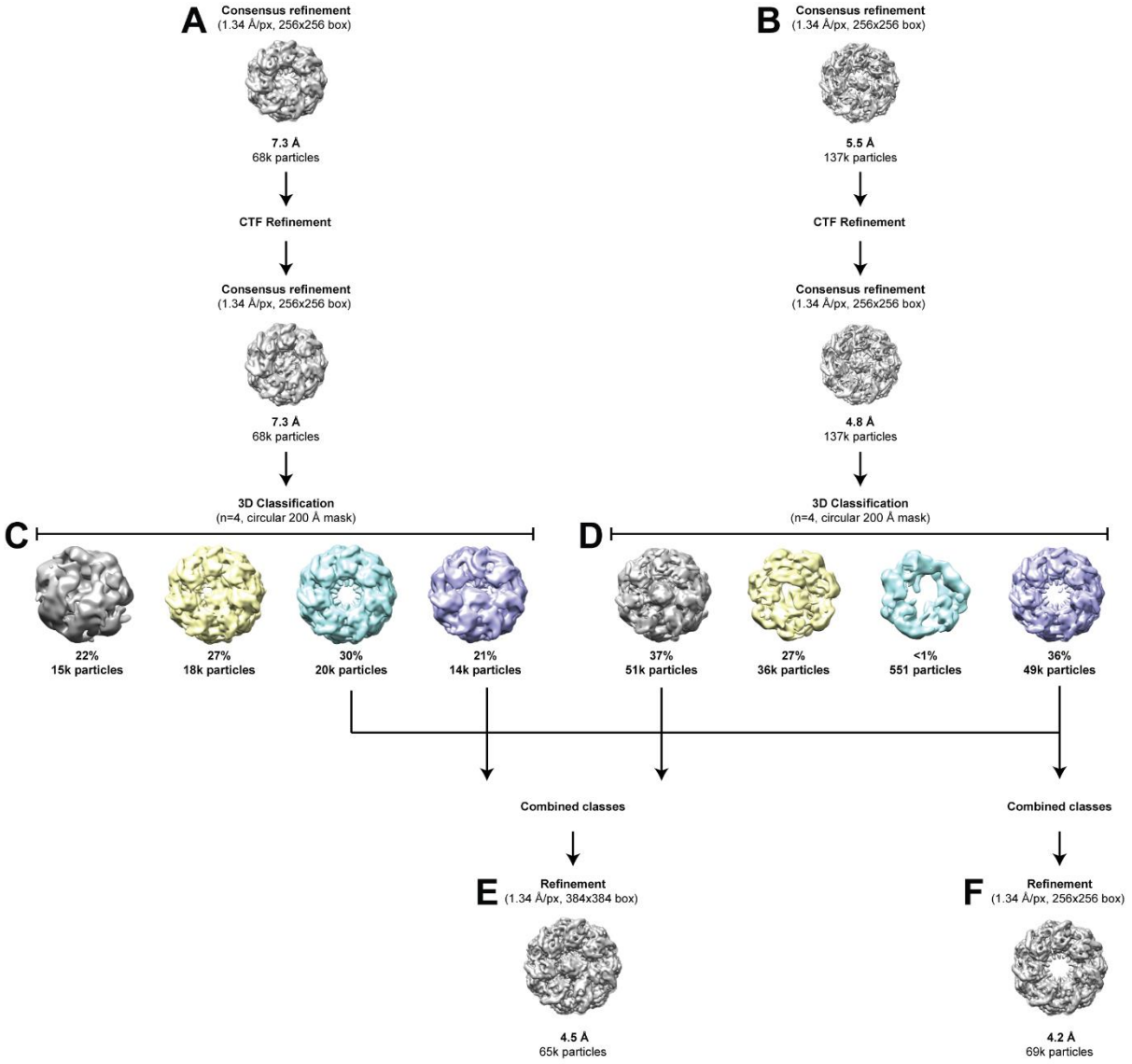

Figure 5-12. Image processing workflow for Chameleon GroEL-Rubisco cryoEM data sets. CryoEM reconstructions from consensus refinements of grid 1 (**A**) and grid 2 (**B**) particles. 3D classification of grid 1 (**C**) and grid 2 (**D**) particles. 3D classification was performed using default Relion settings and four classes. (**E**) Reconstruction from the best substrate-bound 3D class from each Chameleon data set. (**F**) Reconstruction of apo GroEL particles from both Chameleon data sets.

5.4.3. Substrate-bound reconstruction from Chameleon grids

A cryoEM reconstruction of GroEL-Rubisco was calculated from 65k particles (**Fig. 5-13**). The map showed well resolved density for GroEL and strong density for Rubisco (**Fig. 5-13A**, top left and bottom right). The local resolution ranged from 4.2 Å in the equatorial domains, to 14 – 18 Å for Rubisco (**Fig. 5-13A** and **Fig. 5-13B**). Weak density for bound substrate was present in the GroEL *trans* ring (**Fig. 5-13A**, bottom left). Rubisco density was primarily located on one side of the GroEL cavity (**Fig. 5-13B**). Rubisco interaction was relatively strong with three consecutive GroEL subunits (**Fig. 5-13B**, top). A weaker interaction with two additional flanking GroEL subunits was visible at lower threshold (**Fig. 5-13B**, bottom). Most particles had adopted end views (**Fig. 5-13C**). However, enough side and oblique views were present to yield a reconstruction with no obvious signs of anisotropy. The global resolution of the reconstruction was 4.5 Å (**Fig. 5-13D**). We did not observe the extra inter-ring density between Arg-430 and Glu-434. This additional density was only seen in our reconstructions from tilt data (see **Section 5.3.5**).

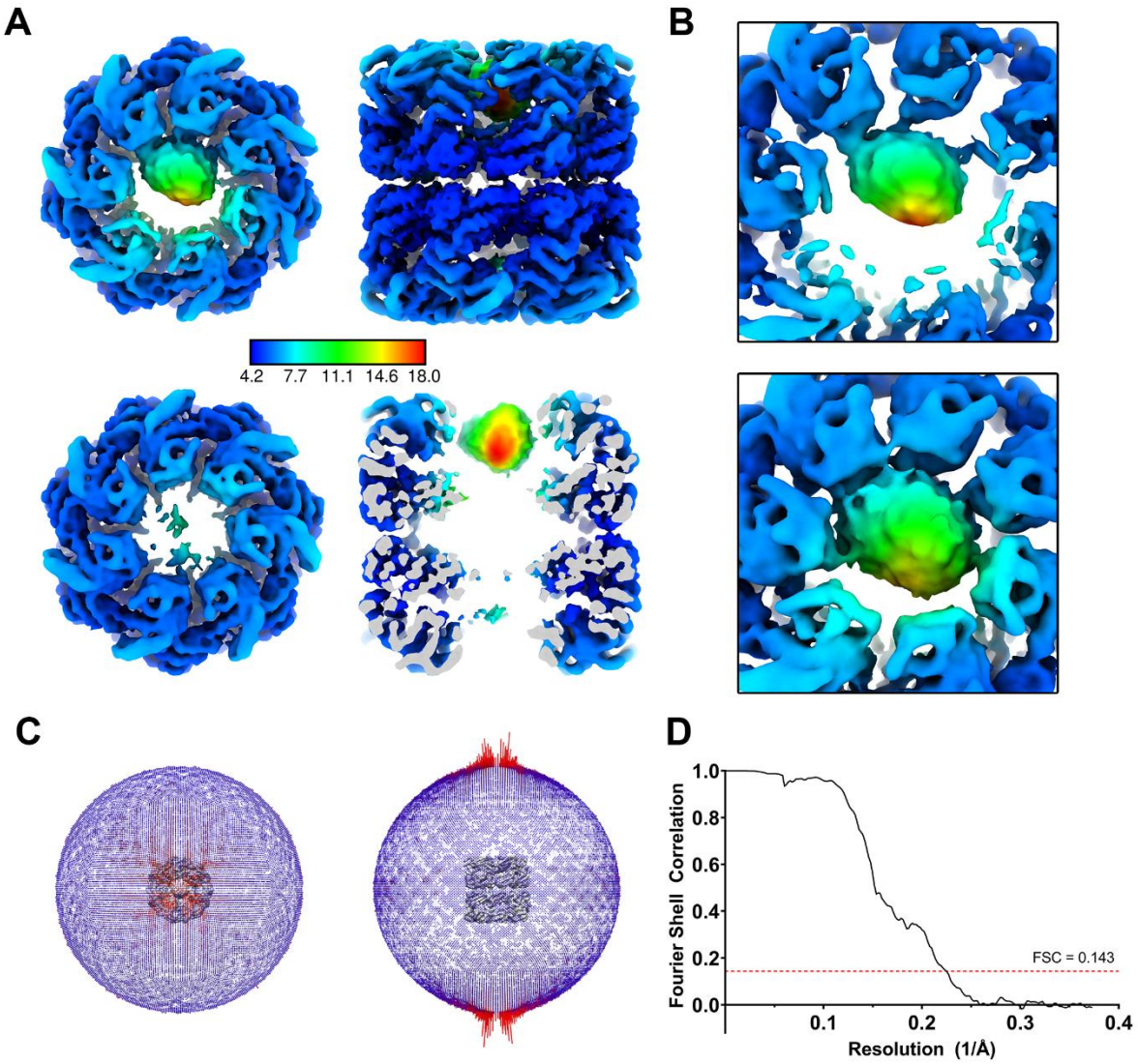

Figure 5-13. CryoEM reconstruction of GroEL-Rubisco from grids prepared using a Chameleon vitrification instrument. (**A**) CryoEM map coloured by local resolution. The colour scale is shown by the central bar, with units in Å. (**B**) Close-up view of the bound non-native Rubisco shown at high threshold (top) and low threshold (bottom) values. (**C**) 3D histogram showing the orientation distribution of GroEL-Rubisco particles. (**D**) Fourier shell correlation (FSC) plot for the calculated half-maps; overall resolution was 4.5 Å.

5.4.4. Mapping the interactions between GroEL and substrate

Non-native Rubisco interacted with multiple GroEL subunits (**Fig. 5-14**). We assigned a number to identify each GroEL subunit (**Fig. 5-14A**). The subunit showing the strongest interaction with

non-native Rubisco was numbered 1. The other six subunits were numbered sequentially in clockwise order when viewing the substrate bound GroEL ring from above (**Fig. 5-14A**). The strongest interactions between GroEL and Rubisco were seen in helix I. In subunit 1, non-native substrate formed strong interactions with Met-267 and Arg-268 (**Fig. 5-14B**). A weaker interaction with subunit 1 Tyr-203 (part of the underlying segment) was resolved at a lower threshold (**Fig. 5-14C**). GroEL subunit 2 showed the same Arg-268 contact but lacked the Met-267 contact to Rubisco (**Fig. 5-14D**). Subunit 2 also showed interactions between Rubisco and the hydrophobic residues Val-264 and Ile-270 (**Fig. 5-14E**). Rubisco interacted weakly with the C-termini of subunit 2 (**Fig. 5-14F**). A second Met-267 in subunit 3 showed a similar interaction to that seen in subunit 1 (**Fig. 5-14G**).

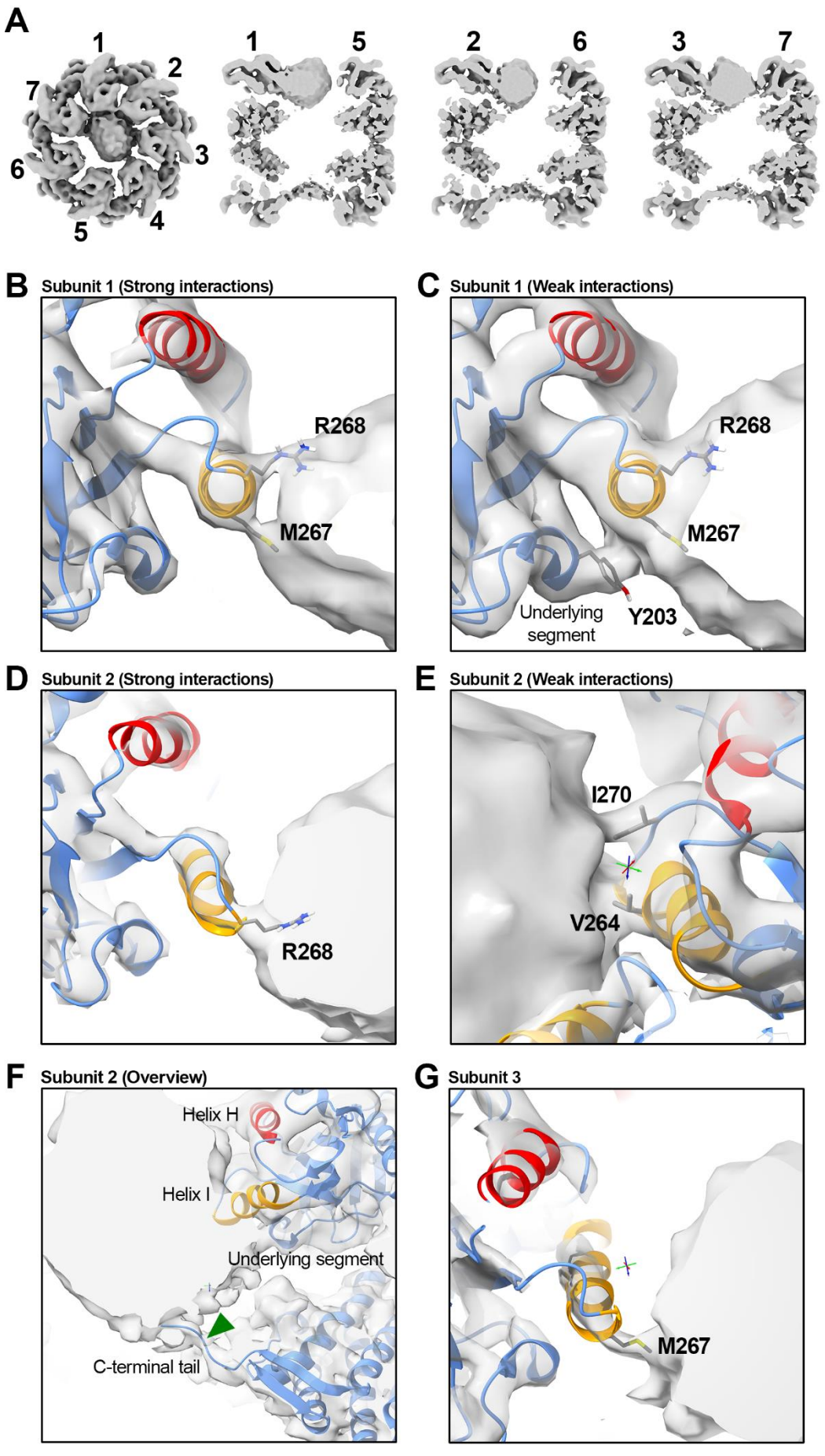
A
1
2
7
3
6
4
5
1
5
2
6
3
7
B
Subunit 1 (Strong interactions)
R268
M267
C
Subunit 1 (Weak interactions)
R268
M267
Underlying segment
Y203
D
Subunit 2 (Strong interactions)
R268
E
Subunit 2 (Weak interactions)
I270
V264
F
Subunit 2 (Overview)
Helix H
Helix I
Underlying segment
C-terminal tail
G
Subunit 3
M267

Figure 5-14. Interactions between GroEL and non-native Rubisco. (**A**) View of the *cis*-cavity of GroEL-Rubisco showing the numbering scheme for GroEL subunits (left). Three central-slice views through the complex showing the extent of binding to GroEL subunits (right). (**B**) View of strong Rubisco contacts in GroEL subunit 1. (**C**) View of weak Rubisco contacts in GroEL subunit 1. (**D**) View of strong Rubisco contacts in GroEL subunit 2. (**E**) View of weak Rubisco contacts in GroEL subunit 2. (**F**) View of the GroEL-Rubisco binding interface of subunit 2. Green arrow highlights contacts made by the disordered C-termini tail. (**G**) View of strong Rubisco contact in GroEL subunit 2. GroEL is depicted as a ribbon diagram and coloured blue. Helices H and I are coloured red and orange.

5.4.5. Non-native Rubisco protrudes from the GroEL cavity

At low density thresholds, we saw non-native Rubisco bound to all seven GroEL subunits. Most residues in helices H and I were involved in contacting substrate. Non-native substrate made contacts with the disordered C-terminal tails of three GroEL subunits. Additionally, weaker density extended ~15 Å above the GroEL *cis* cavity (**Fig. 5-15**). The volume of the GroEL *cis* cavity is approximately half that of the GroEL-GroES cavity and cannot completely encapsulate a 50 kDa Rubisco monomer. Presumably the unbound Rubisco is largely unstructured and mobile. Highly flexible regions of proteins are poorly resolved in cryoEM experiments, leading to weaker or averaged-out density for these features. At very low threshold values, we also observed weak density for bound substrate in the *trans* ring of GroEL (**Fig. 5-15**, middle). This could represent weak binding of non-native substrate, or misclassified/misaligned particles. Rubisco is thought to exert negative cooperativity for substrate binding in the opposite GroEL ring, although the mechanism of this negative cooperativity is poorly understood.

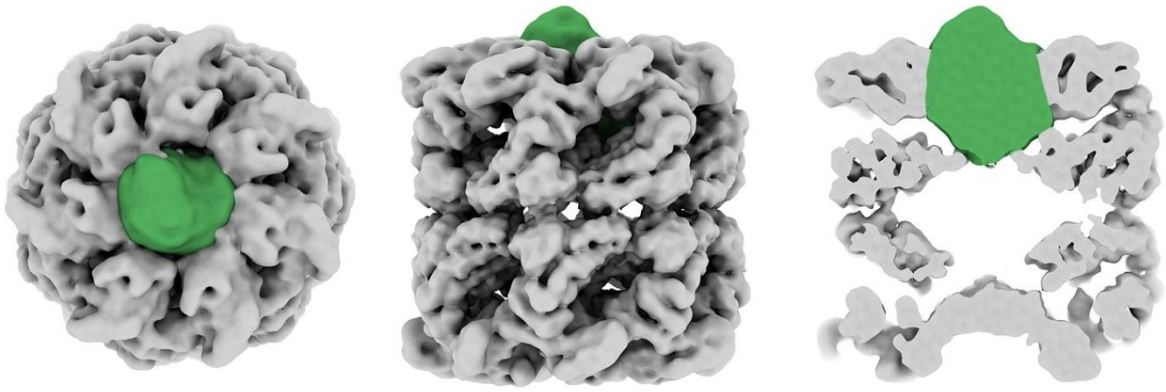

Figure 5-15. Non-native Rubisco protrudes above the GroEL cavity. GroEL-Rubisco viewed at a low density threshold to illustrate weak substrate density. GroEL density is coloured grey, density for bound non-native Rubisco is coloured green.

5.4.6. Estimating the mass of bound non-native Rubisco in cryoEM reconstructions

The volume of a protein can be estimated by its molecular weight, and *vice versa*. A value of 1.21 $Å^3$ per Dalton have been previously calculated (Erickson, 2009). Volume densities (as well as mean average and standard deviation) can be measured in UCSF Chimera (Pettersen et al., 2004), and by dividing the volume by 1.21, we can approximate the molecular mass. We compared the density for bound non-native Rubisco from our two best reconstructions, class 4.2 from the tilt data set (see **Section 5.3.1**) and the final reconstruction from the Chameleon data set (**Fig. 5-13**). Assuming a mass of 851.5 kDa, we set a contour level so that GroEL-Rubisco maps encompassed a volume of ~1,030,000 $Å^3$ (851,500 Da * 1.21). We then isolated the density representing bound Rubisco and measured the volume to estimate its molecular mass. For comparisons, we generated a low-pass filtered volume of a Rubisco monomer from the crystal structure (**Fig. 5-16, left**). We calculated a value of 31.1 kDa for the tilt data set reconstruction, representing ~60% of a Rubisco monomer (**Fig. 5-16**, middle). Substrate density in the Chameleon data set occupied a volume of 60,700 $Å^3$, corresponding to 99% of a Rubisco monomer (**Fig. 5-16**, right). Additionally, we

performed manual fitting of a Rubisco monomer to the substrate density from the Chameleon data set (**Fig. 5-17**).

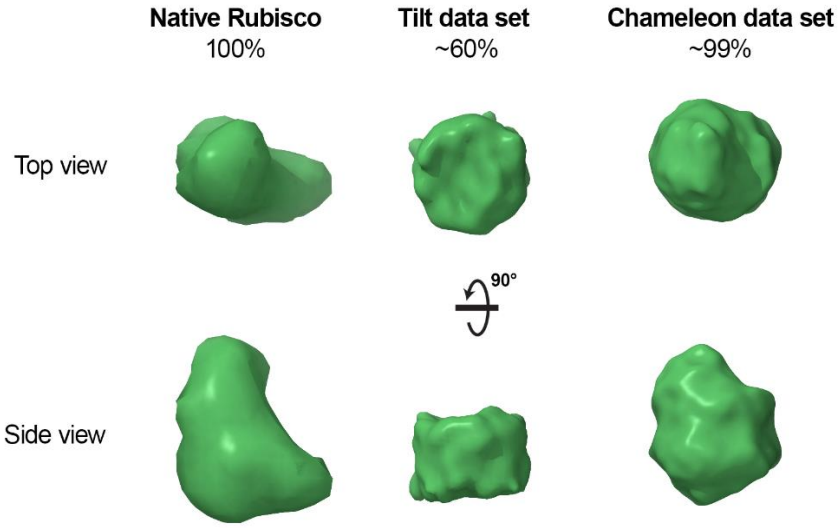

Figure 5-16. Comparison of substrate density from tilt and Chameleon cryoEM data sets. Native Rubisco monomer filtered to 20 Å and contoured to a volume of ~61,000 Å^3 (left). Isolated substrate density from tilt data set class 2 (middle) and Chameleon data set (right).

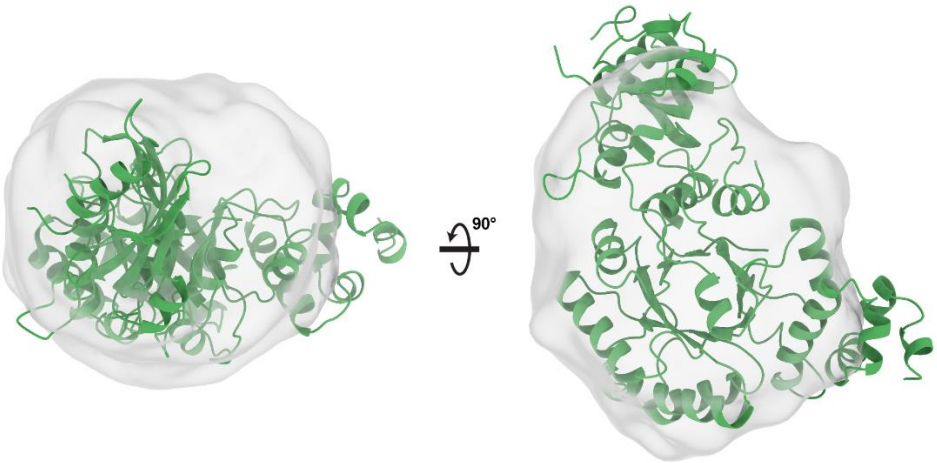

Figure 5-17. Manual fitting of Rubisco monomer. The crystal structure of a Rubisco monomer (PDB: 5RUB) was manually placed inside the isolated substrate density from the final Chameleon data set reconstruction. Estimated mass of the substrate density is 50.2 kDa.

Chapter VI

6. CryoEM study of nucleotide-bound GroEL-Rubisco

We next aimed to determine the effects of ATP binding on GroEL-substrate complexes. In these experiments we formed binary complexes of GroEL bound to non-native Rubisco as previously described (**Chapter V**). We then added ATP or an ATP analogue and prepared the sample for cryoEM. GroEL-ATP has been studied previously and six cryoEM structures of GroELΔD398A-ATP were published in Clare et al. (2012). The 7-fold symmetric GroEL-ATP cryoEM structures characterised the conformational changes that GroEL undergoes upon ATP binding. A later published crystal structure of GroELΔD83A/ΔR197A-ADP, lacking two inter-subunit salt bridges, showed a distinct asymmetry in the GroEL apical domains (Fei et al., 2013). Several different conformations were observed in the crystal structure, and these were in reasonable agreement with the subunit conformations determined by cryoEM in Clare et al. (2012). Importantly, all previously published structures of ATP-bound GroEL were determined in the absence of non-native substrate. The cryoEM structures of GroEL-ATP were also published prior to the recent advances in cryoEM.

Our initial structural efforts using ATP as the nucleotide were hindered by preferred orientation and partial denaturation of GroEL. This was caused by adsorption of the complex to the air-water interface during grid preparation. Despite these technical issues, we saw hints of a novel asymmetric conformation of GroEL-ATP (**Section 6.1**). Several analogues of ATP have proven useful for structural studies of GroEL. We collected preliminary cryoEM data and showed that two ATP analogues trapped GroEL in the same novel asymmetric conformation. Bound substrate was also visible in small subsets of particles. (**Section 6.2**). We then used the Chameleon instrument to reduce the interaction of GroEL with the air-water interface and determine a high resolution cryoEM structure of nucleotide-bound GroEL-Rubisco (**Section 6.3**).

6.1. CryoEM of GroEL-Rubisco-ATP

First, we collected a cryoEM data set of GroEL-Rubisco-ATP. Grids were prepared using a Vitrobot and ATP was added immediately prior to plunge-freezing to minimise dissociation of non-native Rubisco. To mitigate the preferred orientation of the complex we collected images at a tilt angle of 35°.

CryoEM images showed a good distribution of particles in the ice (**Fig. 6-1A**, top). 2D Classification showed a range of views, but most particles were assigned to oblique view class averages of GroEL (**Fig. 6-1A**, bottom). An initial consensus reconstruction of 960k particles exhibited poorly resolved density for intermediate and apical domains of GroEL subunits. The complex appeared flattened on one end (**Fig. 6-1B**, top row, middle). We saw poorly resolved weak density for both rings at low threshold values (**Fib. 6-1B**, bottom row). The refined orientations of particles suggested the same strongly preferred end view, though alleviated by tilting (**Fig. 6-1C**). The initial consensus refinement had an overall Nyquist-limited resolution of 4.4 Å (**Fig. 6-1D**).

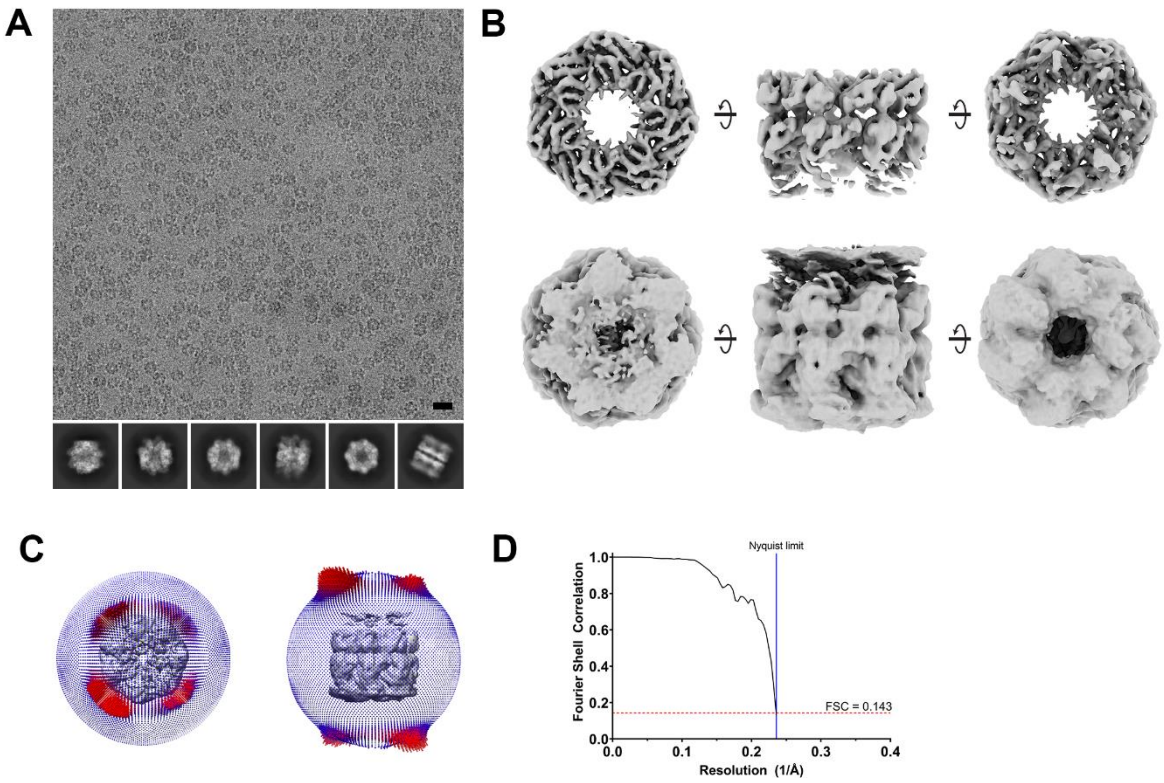

Figure 6-1. Consensus reconstruction of GroEL-Rubisco-ATP. (**A**) Representative motion-corrected cryo micrograph (top; black scale bar = 20 nm) and 2D class averages showing a range of views of the GroEL complex (bottom). (**B**) Reconstruction from 960,480 particles, viewed from the *cis* ring (left), side (middle) and *trans* ring (right) and shown at a high threshold value (top row) and low threshold value (bottom row). (**C**) 3D histogram showing the orientation distribution of GroEL-Rubisco-ATP particles, with the consensus reconstruction shown in the middle of the histogram to illustrate orientation angles. (**D**) Fourier shell correlation (FSC) plot for the calculated half-maps; overall resolution determined by the 0.143 criterion (dashed red line) was 4.4 Å. Blue vertical line represents the Nyquist frequency.

6.1.1. Image processing revealed an asymmetric arrangement of GroEL subunits

We had expected conformational heterogeneity in the data set because GroEL was presumably cycling ATP for several seconds prior to blotting and plunge-freezing. The consensus map showed

that one GroEL ring was always better resolved than the other. We devised an image processing workflow involving 3D classification and particle subtraction and used this to identify subsets of particles in which both GroEL rings were well resolved (**Fig. 6-2**). Initial 3D classification identified two classes of particles in which either the *cis* or *trans* ring of GroEL was well resolved (**Fig. 6-2A**, class 4 and 5). Classes 4 and 5 were refined to 5.3 and 5.2 Å respectively (**Fig. 6-2B** and **Fig. 6-2C**). Next, we subtracted the best-resolved ring from each class and performed 3D classification without alignments on the two sets of subtracted particles (**Fig. 6-2D** and **Fig. 6-2E**). The aim was to identify particles with strong density for both GroEL rings. Several 3D classes showed well-resolved *cis* or *trans* ring density (**Fig. 6-2D**, classes 3 and 4; **Fig. 6-2E**, classes 1 and 2). Particles from these classes were reverted to their original non-subtracted images and refined separately. The best resolved class contained 13k particles and showed the first hints of asymmetry in the *cis* ring. Particles belonging to this class were re-extracted and refined to 5.9 Å (**Fig. 6-2F**). Other 3D classes appeared to show well resolved features, but yielded low resolution reconstructions following 3D refinement (not shown).

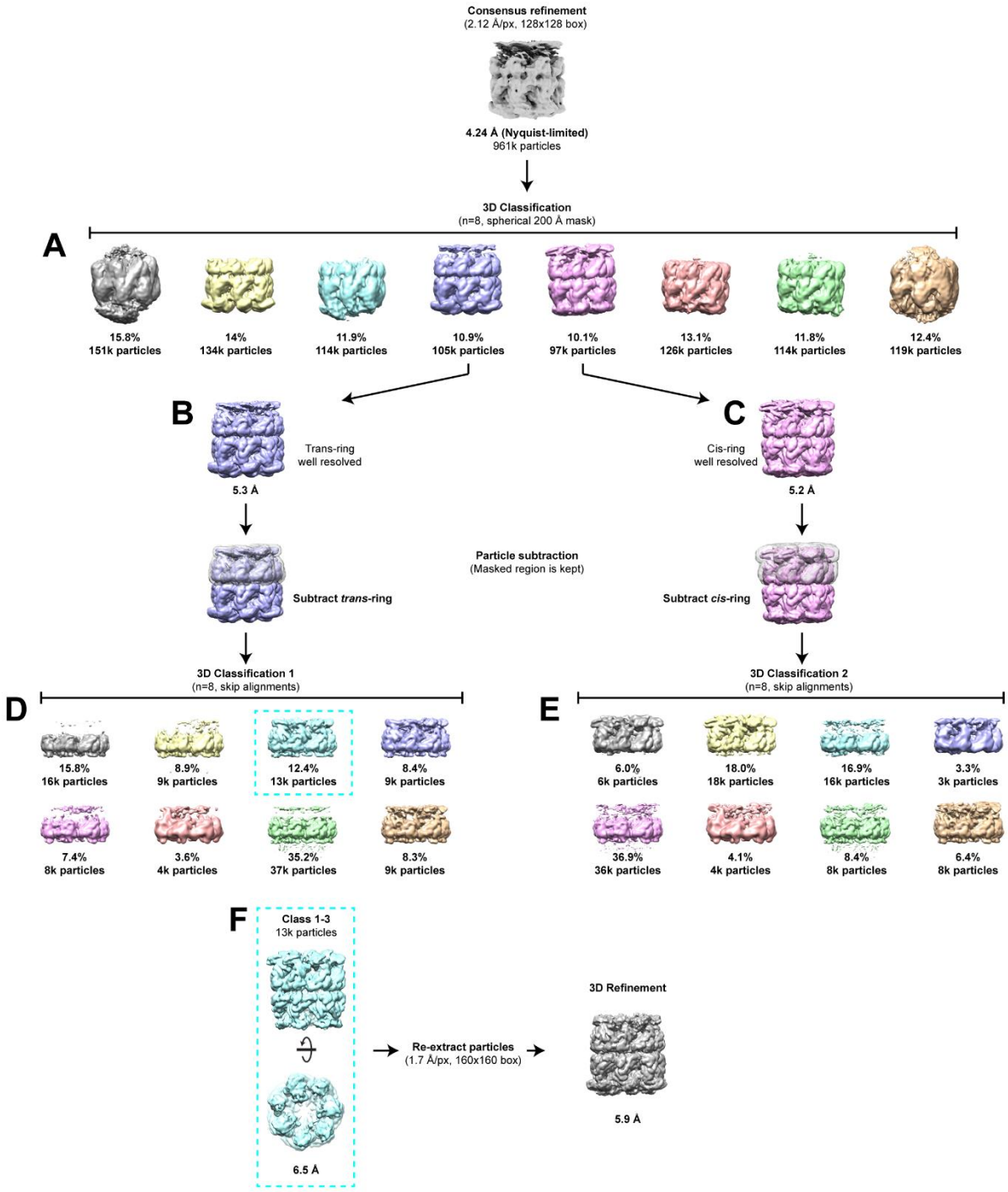

Figure 6-2. Image processing workflow for GroEL-Rubisco-ATP. (**A**) Initial 3D classification. (**B**) Refinement of class 4 and subtraction of the *trans* ring. (**C**) Refinement of class 5 and subtraction of the *cis* ring. (**D**) 3D classification of subtracted class 4 particles. (**E**) 3D classification of subtracted class 5 particles. (**F**) Reconstruction of GroEL-Rubisco-ATP from 13k particles. Resolution was improved to 5.9 Å after re-extraction at a smaller pixel size.

6.1.2. ATP binding induced asymmetry in the *cis* ring of GroEL

The previously published cryoEM reconstructions of GroELΔD398A-ATP were refined with C7 symmetry; any intra-ring asymmetry would not have been resolved. The authors of Clare et al. (2012) observed mild asymmetry when performing asymmetric refinements of some classes. Notably, an asymmetric refinement of their Rs2 class suggested a raised conformation for 1 – 2 GroEL subunits. Presumably there were too few particles to calculate moderate-resolution asymmetric reconstructions.

Here, we performed asymmetric 3D refinement and calculated a 5.9 Å map of GroEL-Rubisco-ATP from 13k particles (**Fig. 6-3**). The local resolution of the map varied from 4.8 Å in the equatorial domains to 8 Å or lower in the apical domains (**Fig. 6-3A**). GroEL subunits appeared to adopt one of several conformations. We used rigid body fitting of GroELΔD398A-ATP subunits to assign a conformation to each subunit.

GroEL subunits in the *cis* ring were arranged in a novel set of conformations. We numbered GroEL subunits sequentially in a clockwise order (**Fig. 6-3B**). First, a triad of non-consecutive subunits (GroEL subunits 1, 3 and 6) showed an Rs1-like conformation (**Fig. 6-3B**, top left; **Fig. 6-3E**, left). In the Rs1 state, the intermediate and apical domains have undergone a 35° sideways tilt as a single rigid body compared to apo GroEL (Clare et al., 2012). The triad of Rs1 subunits had well-resolved density for their apical domains, including hydrophobic helices H and I (**Fig. 6-3E**, left). Next, two consecutive subunits (GroEL subunits 4 and 5) showed a conformation in which helices H and I were packed closely together. We fitted the Rs1 conformation to these two subunits. However, the fit was relatively poor and apical domains clashed (**Fig. 6-3B**, top right). The apical domain density of subunits 4 and 5 was weaker compared to subunits 1, 3 and 6, suggesting an ensemble of states. Continuous density for helices H and I of both subunits was visible at a low threshold value (**Fig. 6-3E**, middle). The remaining two non-consecutive subunits (GroEL subunits 2 and 7) were the least well resolved (**Fig. 6-3B**, bottom left). In these subunits,

parts of the apical domains, including helices H and I, were not resolved and were absent from the reconstruction. As a result, we did not model these residues (**Fig. 6-3E**, right). It is likely that these regions were either disordered or had been denatured following adsorption at the air-water interface. Lastly, the *trans* ring of GroEL-Rubisco-ATP displayed approximate 7-fold symmetry. *Trans* ring subunits had adopted an Rd3-like conformation (**Fig. 6-3A**, bottom left; **Fig. 6-3B**, bottom right).

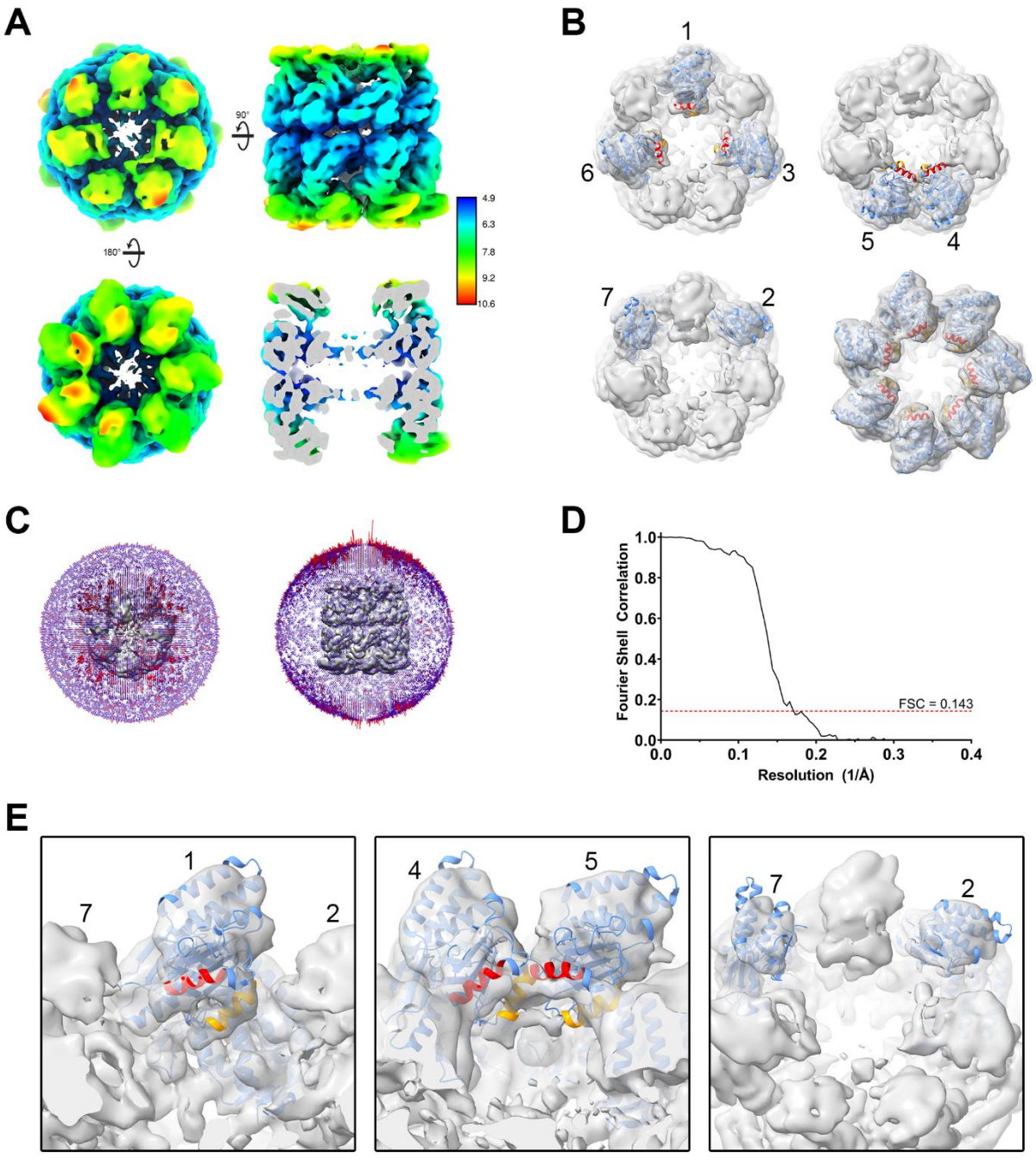

Figure 6-3. CryoEM structure of GroEL-Rubisco-ATP. (**A**) Locally filtered reconstruction coloured by the displayed colour key (right) showing a view from the *cis* ring (top left), the *trans* ring (bottom left), the side (top right) and a central slice through the complex (bottom right). (**B**) Rigged body fitting of GroELΔD398A-ATP subunits; Rs1 conformation (PDB: 4AAQ) fitted to subunits 1, 3 – 6. The Rs1 conformation is also fitted to subunits 2 and 7; but residue ranges 226 – 277, 301 – 316, and 199 – 212 were not modelled. (**C**) 3D histogram showing the orientation distribution of GroEL-Rubisco-ATP particles, with the locally filtered reconstruction included to illustrate orientation angles. (**D**) Fourier shell correlation (FSC) plot for the calculated half-maps; overall resolution determined by the 0.143 criterion (dashed red line) was 5.9 Å. (**E**) Views showing the quality of fit for subunit 1 (left), subunits 4 and 5 (middle) and subunits 2 and 7 (right).

6.1.3. Nucleotide was bound in all 7 subunits in both GroEL rings

Bound nucleotide was visible as additional density near the N-terminus of helix D (**Fig. 6-4**). For both the *cis* and *trans* ring subunits, density was strongest for nucleotide phosphates. Phosphate density was marginally stronger in the *cis* ring, suggesting that our structure represented GroEL-ATP_7-ADP_7. However, we could not reliably distinguish between ADP and ATP. Density was also relatively weak for the surrounding catalytic residues D52, D87 and D398.

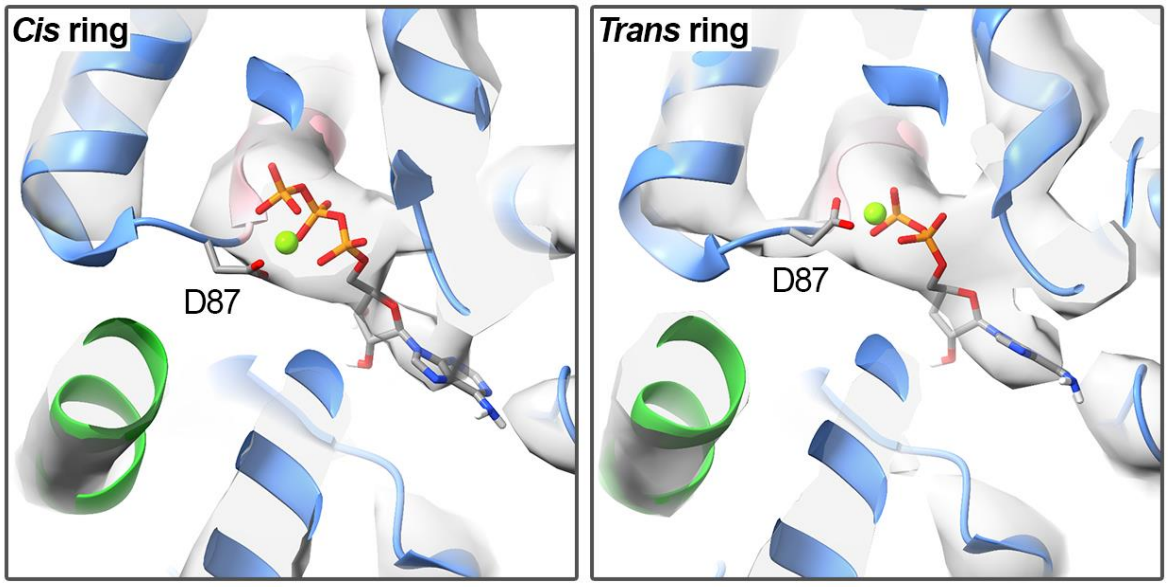

Figure 6-4. Nucleotide density in the *cis* and *trans* rings of GroEL-Rubisco-ATP. GroEL model is represented by a ribbon diagram and coloured blue. Nucleotide and D87 residue shown as a stick representation. Helix D (aa 88 – 110) is coloured pink. Helix M (aa 386 – 408) is coloured green.

6.1.4. GroEL-Rubisco-ATP was prone to denaturation at the air-water interface

Our reconstruction of GroEL-ATP was calculated from 13k particles; around 1.5% of the total number of particles after 2D classification. During cryoEM image processing it is typical to discard objects that do not represent the complex of interest. These are usually high contrast features that are mis-identified as particles during particle picking. Some common examples are ice contaminants, edges of carbon or graphene, and aggregated protein. These non-particles are filtered out and discarded during 2D and 3D classification. However, most class averages in our initial 3D classification did represent GroEL-ATP particles (see **Fig. 6-2A**). Discarded classes showed clear density for GroEL equatorial and intermediate domains. This suggested that the GroEL-ATP particles were at least partially intact but had been damaged. We calculated a reconstruction from the most populated "bad" 3D class to understand the cause of this missing density (**Fig. 6-5**). The resulting cryoEM map exhibited relatively well resolved equatorial domains with recognisable density for α-helices (**Fig. 6-5A**). Apical and intermediate domains were poorly resolved, an effect that was noticeably more prominent in the top ring compared to the bottom ring (**Fig. 6-5A**, right). At a low density threshold value, the density for the top ring was incomplete and appeared smeared and flattened (**Fig. 6-5B**). We attributed this to partial denaturation of GroEL at the air-water interface prior to plunge-freezing (crudely illustrated in **Fig. 6-5C**).

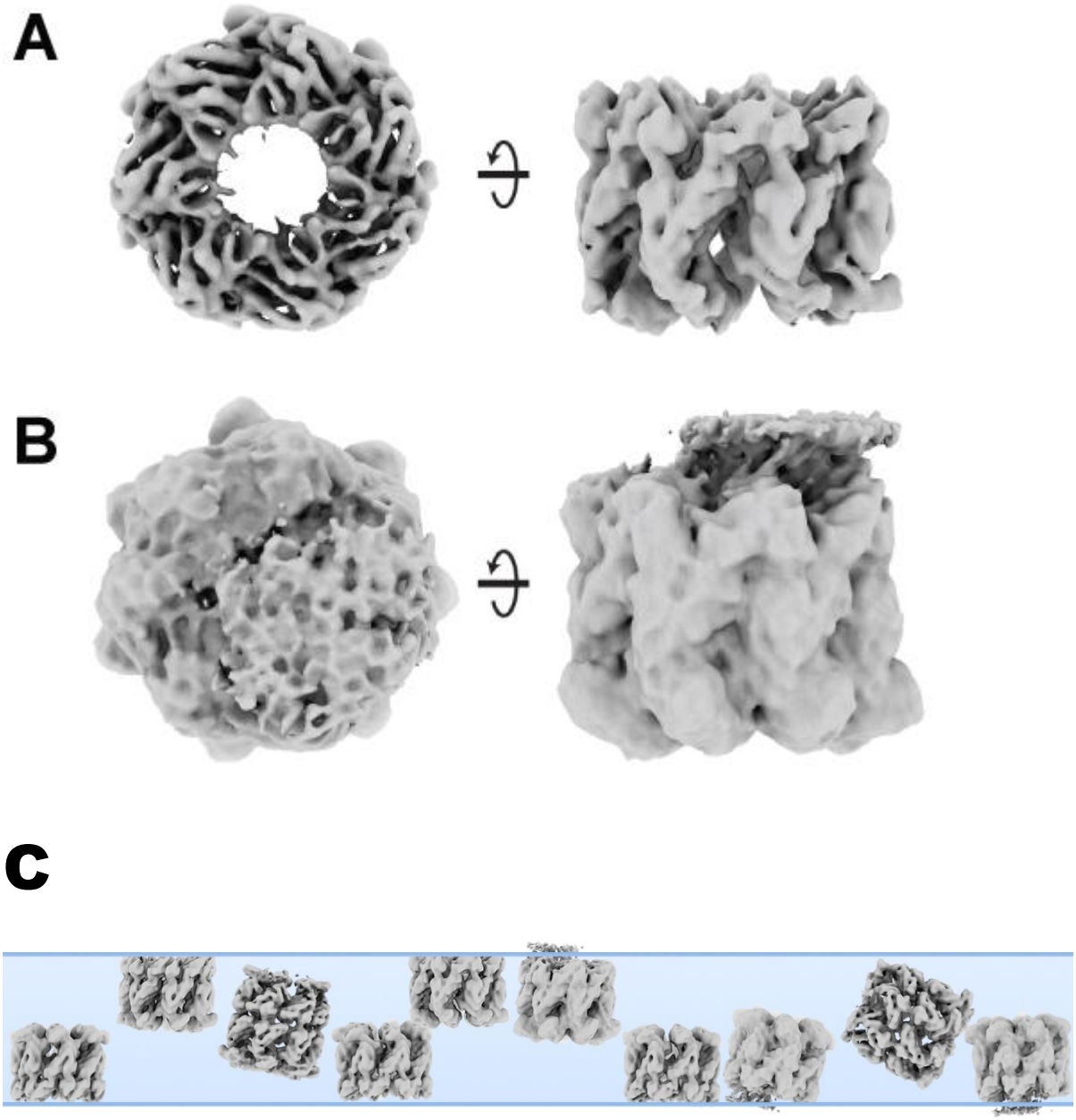

Figure 6-5. Partial denaturation of GroEL-Rubisco-ATP at the air-water interface. (**A**) 3D Reconstruction of damaged GroEL-Rubisco-ATP from 134k particles. (**B**) Reconstruction shown at a low threshold value to highlight the unresolved apical domain density of one GroEL ring (**C**) Illustration of intact and damaged GroEL-Rubisco-ATP particles embedded in the thin ice layer (blue) of a frozen EM grid.

6.1.5. Protein denaturation during cryoEM grid preparation

Proteins that adsorb to an air-water interface can partially or completely denature, as has been demonstrated by electron cryo-tomography of the yeast protein fatty acid synthase (D'Imprima et al., 2019). We observed a similar effect for GroEL-Rubisco-ATP.

GroEL forms a stable tetradecameric complex that can withstand high temperatures (> 65 °C) and high concentrations of chemical denaturant (> 1 M urea) without adverse effects. For example, a late step of the GroEL purification protocol uses 20% (v/v) methanol to strip out contaminating non-native substrate proteins. Yet this does not affect the GroEL structure. We did not observe denaturation of nucleotide-free GroEL in our earlier cryoEM experiments (see **Chapter V**). ATP-bound GroEL is likely more susceptible to this kind of denaturation. The ATP-induced conformational changes of GroEL involve the breaking and forming of salt bridges between adjacent apical domains (Clare et al., 2012). Individual GroEL subunits adopt more open states once these salt bridges are broken. These apical domains are likely prone to denaturation. In nucleotide-free GroEL-Rubisco, the GroEL subunits are therefore protected against denaturation by these inter-subunit contacts.

6.1.6. GroEL-Rubisco-ATP was not an ideal sample for cryoEM

In summary, GroEL-Rubisco-ATP complexes suffered from a strongly preferred orientation and were prone to partial denaturation at the air-water interface. The best reconstruction hinted at an asymmetric structure of GroEL. However, resolution was limited and did not permit confident atomic modelling. We decided not to pursue a cryoEM structure of GroEL-Rubisco-ATP. Instead, we opted to use an ATP analogue in the hope of trapping the asymmetric conformation.

6.2. Preliminary CryoEM of GroEL-Rubisco with ATP analogues

The ADP-metal complexes ADP·BeF_x, ADP·AlF_x and ADP·VO_4 have been used to mimic the γ-phosphate of ATP along the reaction coordinate of ATP hydrolysis (Chaudhry et al., 2003). Specifically, ADP·BeF_x is a ground state analogue, ADP·AlF_x is a transition-state analogue of ATP hydrolysis, and ADP·VO_4 is a post-hydrolysis analogue (Chaudhry et al., 2003). GroEL-GroES complexes formed with either ADP·BeF_x or ADP·AlF_x can trigger productive folding of Rhodanese and Malate Dehydrogenase, two stringent substrates of GroEL. The post-hydrolysis analogue ADP·VO_4, or ADP alone, both fail to trigger folding of the same substrates (Chaudhry et al., 2003). Additionally, ADP·BeF_x and ADP·AlF_x have been used in biochemical and structural studies (cryoEM and X-ray crystallography) of both GroEL and the eukaryotic chaperonin TRiC (e.g. in Balchin et al., 2018, Chaudhry et al., 2003 and Clare et al., 2009).

We prepared complexes of GroEL-Rubisco-ADP·BeF_x and GroEL-Rubisco-ADP·AlF_x. Our earlier native PAGE experiments examined these complexes and showed no differences in electrophoretic migration or substrate occupancy between the two (see **Section 4.1**). We performed negative stain EM and low resolution cryoEM of both complexes and again, observed no obvious structural differences (not shown).

6.2.1. CryoEM of GroEL-Rubisco with two different ATP analogues

We collected preliminary high resolution cryoEM data sets (< 2,000 micrograph movies each) of GroEL-Rubisco-ADP·BeF_x and GroEL-Rubisco-ADP·AlF_x. The aims of these small data sets were several-fold: i) to confirm that one or both ATP analogues trap the asymmetric complex observed in our reconstruction of GroEL-Rubisco-ATP, ii) to explore any structural differences between the two complexes, and iii) to decide which complex we would ultimately take forward

and use in cryoEM grid preparation with the Chameleon instrument. Our Chameleon sessions were limited (with a 3-month lead time) and it was important to identify the most promising sample.

6.2.2. Preliminary CryoEM of GroEL-Rubisco-ADP·BeF_x and GroEL-Rubisco-ADP·BeF_x

For preliminary data sets, cryoEM grids of GroEL-Rubisco-ADP·BeF_x and GroEL-Rubisco-ADP·AlF_x were prepared using a Vitrobot. Micrograph movies were collected of GroEL-Rubisco-ADP·BeF_x and GroEL-Rubisco-ADP·AlF_x (1,770 and 1,930 micrograph movies respectively). CryoEM images was collected data at a 30° tilt to mitigate the preferred orientation of GroEL-substrate and GroEL-nucleotide complexes.

Overall, both data sets yielded similar results (**Fig. 6-6** and **Fig. 6-7**). CryoEM images showed a good distribution of particles in the ice (**Fig. 6-6A**, top; **Fig. 6-7A**, top). 2D Classification showed clear secondary structure elements. Class averages represented a range of views of GroEL, including oblique and side views (**Fig. 6-6A**, bottom; **6-7A**, bottom). Initial consensus reconstructions showed generally well resolved asymmetric GroEL complexes (**Fig. 6-6B** and **Fig. 6-7B**). Clear density for α-helices was visible in the equatorial domains of GroEL subunits. Although the density for GroEL apical domains was weak, denaturation appeared less severe compared to the ATP-bound complex. Most particles represented oblique views of GroEL (**Fig. 6-6C** and **Fig. 6-7C**). For initial image processing steps, particles were extracted at 2.1 Å/px and the resolution of consensus reconstructions were Nyquist-limited at 4.2 Å (**Fig. 6-6D** and **Fig. 6-7D**). Down sampling (or binning) sped up computations and was typical for our early/intermediate image processing stages.

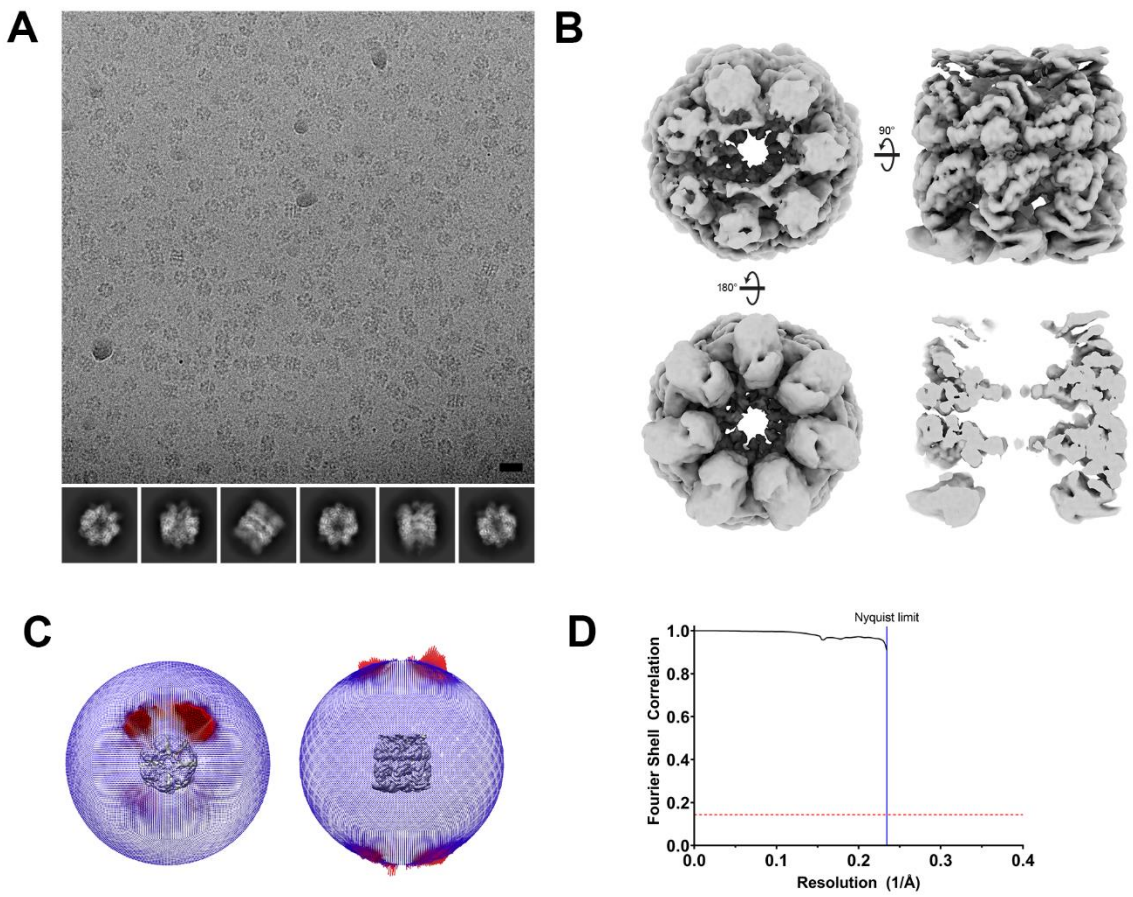

Figure 6-6. Consensus reconstruction of GroEL-Rubisco-ADP·BeF$_x$. (**A**) Representative motion-corrected cryoEM micrograph (top; black scale bar = 20 nm) and six 2D class averages showing a range of views of the GroEL complex (bottom). (**B**) Reconstruction from 666,080 particles, viewed from the *cis* ring (top left), side (top right) and *trans* ring (bottom left) and a central slice (bottom right). (**C**) 3D histogram showing the orientation distribution of GroEL-Rubisco-ADP·BeF$_x$. (**D**) Fourier shell correlation (FSC) plot for the calculated half-maps; overall resolution was Nyquist-limited at 4.2 Å.

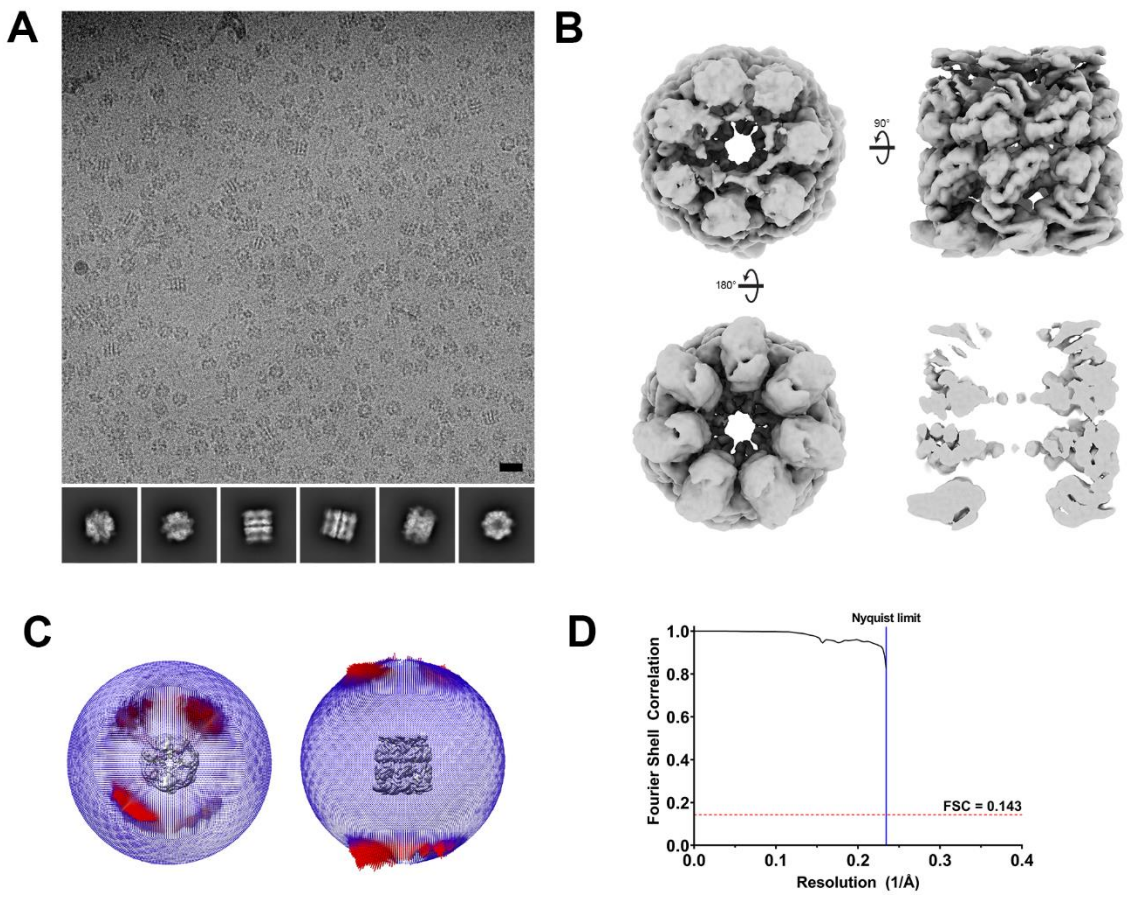

Figure 6-7. Consensus reconstruction of GroEL-Rubisco-ADP·AlF$_x$. (**A**) Representative motion-corrected cryoEM micrograph (top; black scale bar = 20 nm) and six 2D class averages showing a range of views of the GroEL complex (bottom). (**B**) Reconstruction from 812,800 particles, viewed from the *cis* ring (top left), side (top right) and *trans* ring (bottom left) and a central slice (bottom right). (**C**) 3D histogram showing the orientation distribution of GroEL-Rubisco-ADP·BeF$_x$. (**D**) Fourier shell correlation (FSC) plot for the calculated half-maps; overall resolution was Nyquist-limited at 4.2 Å.

6.2.3. Image processing of GroEL-Rubisco-ADP·BeF$_x$/ADP·AlF$_x$ cryoEM data sets

We used similar image processing workflows for GroEL-Rubisco-ADP·BeF$_x$ and GroEL-Rubisco-ADP·AlF$_x$ data sets (**Fig. 6-8** and **Fig. 6-9**). Particles from 2D classification were

extracted at 2.13 Å/px and refined (**Fig. 6-8A** and **Fig. 6-9A**). Because cryoEM images were collected at a high tilt angle, we performed CTF refinement and 3D refinement prior to 3D classification (**Fig. 6-8B** and **Fig. 6-9B**). This ensured more accurate per-particle defocus values. Both data sets yielded 4.3 Å (Nyquist-limited) consensus maps at this stage (**Fig. 6-8B** and **Fig. 6-9B**). Particles were subjected to 3D classification using a 200 Å circular mask to remove damaged or "junk" particles (**Fig. 6-8C** and **Fig. 6-9C**). Denaturation at the air-water interface was apparent for some classes (e.g. **Fig. 6-8C**, class 1 and **Fig. 6-9C**, class 4). Particles from the best-resolved classes were grouped together and refined (**Fig. 6-8D** and **Fig. 6-9D**). We then used 3D classification without alignments (**Fig. 6-8E** and **Fig. 6-9E**). Several well resolved 3D classes showed the same asymmetric state seen for GroEL-Rubisco-ATP. These classes were combined and refined to 4.3 Å (**Fig. 6-8F** and **Fig. 6-9F**). Focussed 3D classification with a mask encompassing the GroEL *cis* cavity allowed us to identify particles with strong substrate density (**Fig. 6-8G** and **Fig. 6-9G**). Particles from the best 3D class were re-extracted at 1.6 Å/px and refined (**Fig. 6-8H** and **Fig. 6-9H**).

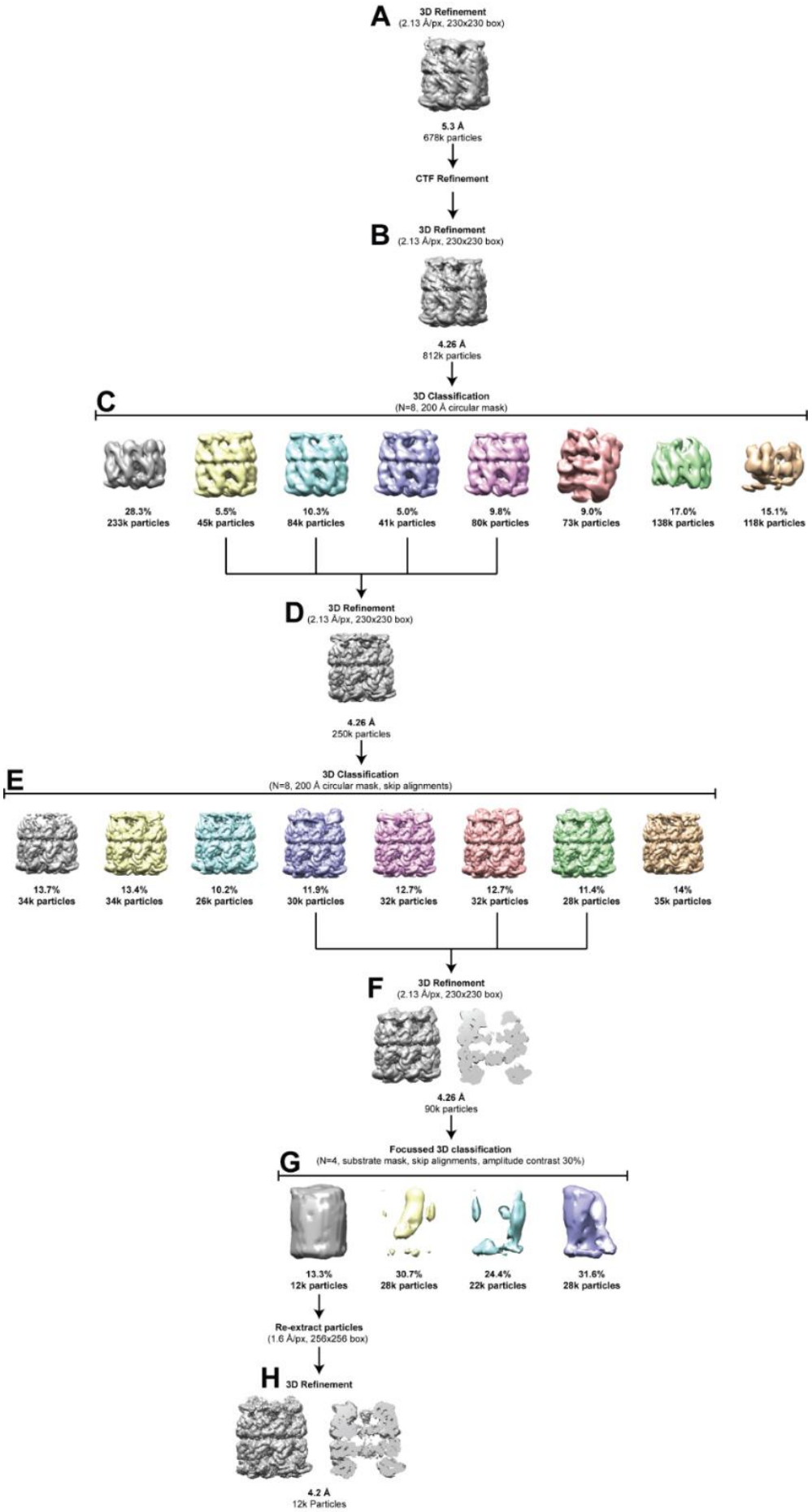
A
3D Refinement
(2.13 Å/px, 230x230 box)
5.3 Å
678k particles
CTF Refinement
B
3D Refinement
(2.13 Å/px, 230x230 box)
4.26 Å
812k particles
C
3D Classification
(N=8, 200 Å circular mask)
28.3%
233k particles
5.5%
45k particles
10.3%
84k particles
5.0%
41k particles
9.8%
80k particles
9.0%
73k particles
17.0%
138k particles
15.1%
118k particles
D
3D Refinement
(2.13 Å/px, 230x230 box)
4.26 Å
250k particles
E
3D Classification
(N=8, 200 Å circular mask, skip alignments)
13.7%
34k particles
13.4%
34k particles
10.2%
26k particles
11.9%
30k particles
12.7%
32k particles
12.7%
32k particles
11.4%
28k particles
14%
35k particles
F
3D Refinement
(2.13 Å/px, 230x230 box)
4.26 Å
90k particles
G
Focussed 3D classification
(N=4, substrate mask, skip alignments, amplitude contrast 30%)
13.3%
12k particles
30.7%
28k particles
24.4%
22k particles
31.6%
28k particles
Re-extract particles
(1.6 Å/px, 256x256 box)
H
3D Refinement
4.2 Å
12k Particles

Figure 6-8. Image processing of GroEL-Rubisco-ADP·BeF$_x$. (**A**) Consensus reconstruction. (**B**) Consensus reconstruction following CTF refinement (**C**) 3D classification using alignments. (**D**) Reconstruction from particles in the best 3D classes. (E**)** 3D classification without alignments. (**F**) Reconstruction from particles in the best 3D classes. (**G**) Focussed 3D classification with a mask encompassing the *cis* cavity of GroEL. Alignments were skipped and amplitude contrast was set to 0.3. (**H**) Reconstruction of GroEL-Rubisco-ADP·BeFx.

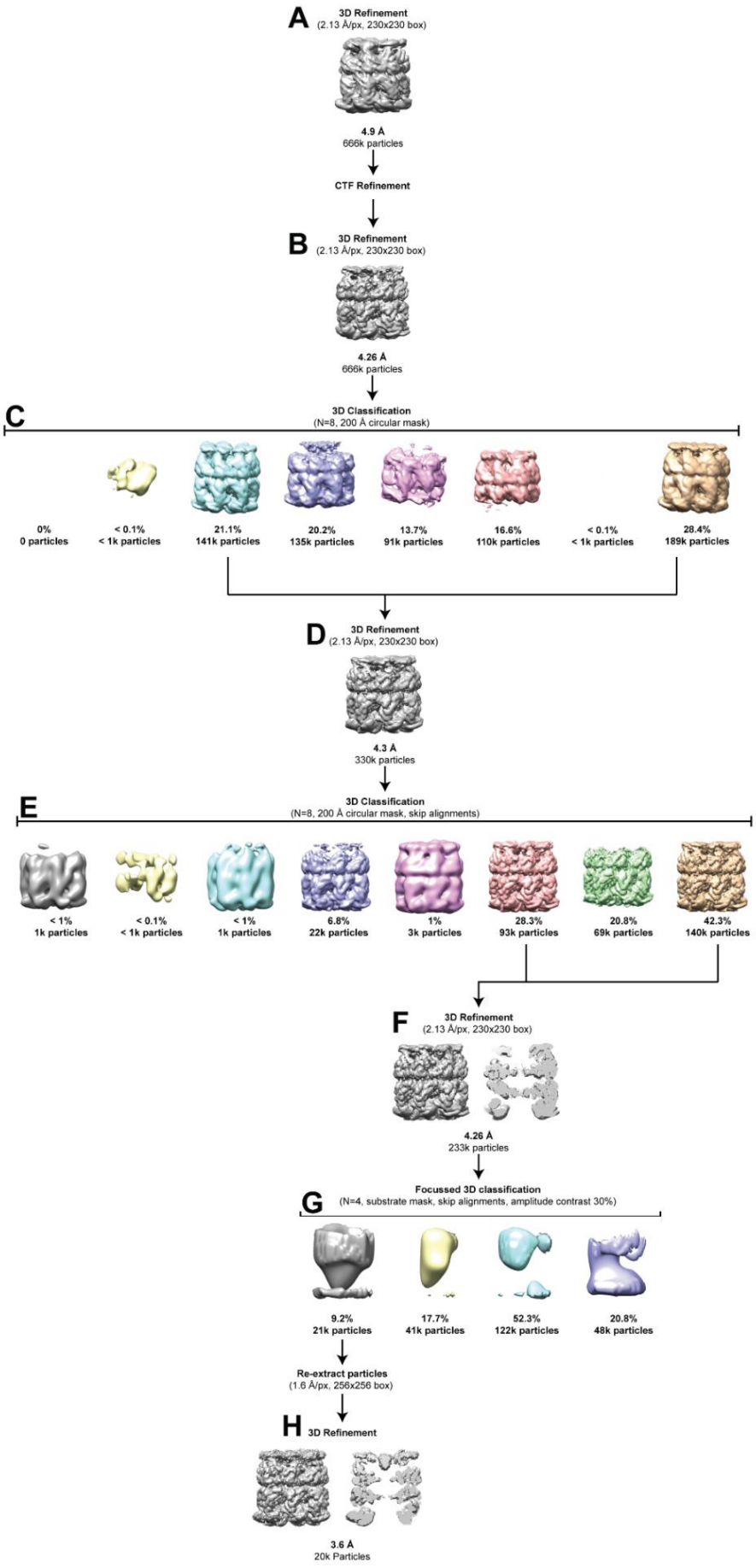
A
3D Refinement
(2.13 Å/px, 230x230 box)
4.9 Å
666k particles
CTF Refinement
B
3D Refinement
(2.13 Å/px, 230x230 box)
4.26 Å
666k particles
C
3D Classification
(N=8, 200 Å circular mask)
0%
0 particles
< 0.1%
< 1k particles
21.1%
141k particles
20.2%
135k particles
13.7%
91k particles
16.6%
110k particles
< 0.1%
< 1k particles
28.4%
189k particles
D
3D Refinement
(2.13 Å/px, 230x230 box)
4.3 Å
330k particles
E
3D Classification
(N=8, 200 Å circular mask, skip alignments)
< 1%
1k particles
< 0.1%
< 1k particles
< 1%
1k particles
6.8%
22k particles
1%
3k particles
28.3%
93k particles
20.8%
69k particles
42.3%
140k particles
F
3D Refinement
(2.13 Å/px, 230x230 box)
4.26 Å
233k particles
Focussed 3D classification
G
(N=4, substrate mask, skip alignments, amplitude contrast 30%)
9.2%
21k particles
17.7%
41k particles
52.3%
122k particles
20.8%
48k particles
Re-extract particles
(1.6 Å/px, 256x256 box)
H
3D Refinement
3.6 Å
20k Particles

Figure 6-9. Image processing of GroEL-Rubisco-ADP·AlF$_x$. (**A**) Consensus reconstruction. (**B**) Consensus reconstruction following CTF refinement (**C**) 3D classification using alignments. (**D**) Reconstruction from particles in the best 3D classes. (E**)** 3D classification without alignments. (**F**) Reconstruction from particles in the best 3D classes. (**G**) Focussed 3D classification with a mask encompassing the *cis* cavity of GroEL. Alignments were skipped and amplitude contrast was set to 0.3. (**H**) Reconstruction of GroEL-Rubisco-ADP·AlF$_x$.

6.2.4. First views of a non-native substrate bound to GroEL-ADP·BeF$_x$/AlF$_x$

We used a focussed 3D classification approach to identify subsets of particles with strong substrate density from both data sets. From these classes we calculated reconstructions of GroEL-Rubisco-ADP·BeF$_x$ from 12k particles, and GroEL-Rubisco-ADP·AlF$_x$ from 20k particles (**Fig. 6-10**). Both reconstructions showed an asymmetric conformation of the GroEL *cis* ring. The reconstruction was similar to our low-resolution reconstruction of GroEL-Rubisco-ATP (see **Section 6.1.2**). This confirmed that ADP·BeF$_x$ and ADP·AlF$_x$ were able to mimic the action of ATP. The overall resolution of the ADP·BeF$_x$-bound complex was 3.8 Å. Local resolution ranged from 3.8 Å in the equatorial domains, to 7 – 11 Å in the apical domains (**Fig. 6-10A**). The overall resolution of the ADP·AlF$_x$-bound complex was slightly higher, ranging from 3.2 Å in the equatorial domains, to 7 – 8 Å in the apical domains (**Fig. 6-10C**). The ADP·BeF$_x$-bound complex showed extensions rising from several GroEL subunits (**Fig. 6-10A**, black arrowheads). We attributed these to the apical domains of GroEL subunits 2, 5 and 7. These had been unresolved and left unmodelled in our earlier reconstruction of GroEL-ATP (see **Fig. 6-2**). GroEL apical domains typically face toward the centre of the cavity in ATP-bound states. However, the apical domains of GroEL subunits 2, 5 and 7 instead appeared to face upward. This conformation of GroEL was reminiscent of the GroES-bound state and has not previously been observed in the

absence of GroES (see **Fig. 1-8** in **Chapter I** for a comparison of apo, ATP-bound and GroES-bound subunits of GroEL).

At high threshold values, the interactions between GroEL and substrate were not well resolved in either map (**Fig. 6-10B** and **Fig. 6-10D**, left; at this threshold, the density between GroEL and non-native substrate is not continuous). At lower density thresholds we observed interactions between non-native Rubisco and GroEL helix I, the lower of the two hydrophobic helices (**Fig. 6-10B** and **Fig. 6-10D**, right). This interaction appeared strongest in GroEL subunit 1. We also saw hints of substrate interaction with helix H and the disordered C-termini of GroEL. However, these interactions were poorly resolved, and we could not confidently model interactions involving individual GroEL residues. Substrate also appeared to be more confined to the GroEL cavity, and we saw no extra density protruding as we did for our GroEL-Rubisco cryoEM reconstructions from Chameleon grids (see **Section 5.4.5**).

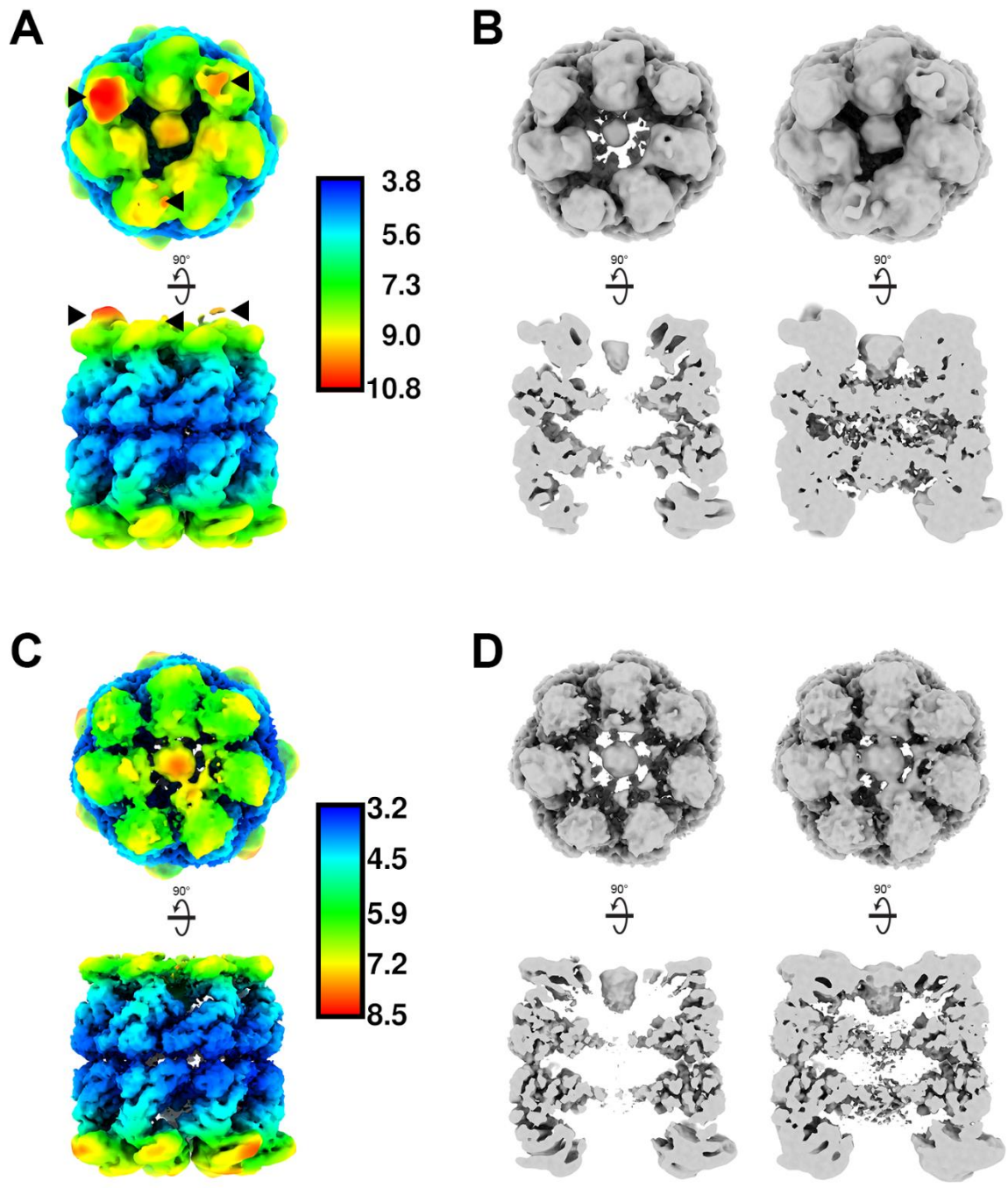

Figure 6-9. CryoEM reconstructions of GroEL-Rubisco-ADP·BeF_x and GroEL-Rubisco-ADP·AlF_x from preliminary data sets. Local resolution of (**A**) GroEL-Rubisco-ADP·BeF_x and (**B**) GroEL-Rubisco-ADP·AlF_x, keys shown on the right with units of Å. *Cis* ring and central side view slices of (**C**) GroEL-Rubisco-ADP·BeF_x and (**D**) GroEL-Rubisco-ADP·AlF_x shown at a high (left) and low (right) density threshold.

6.2.5. Structures of GroEL-Rubisco-ADP·BeF_x/AlF_x

To determine molecular models of GroEL-Rubisco-ADP·BeF_x and GroEL-Rubisco-ADP·AlF_x, we rigid-body fitted the previously published model of the GroEL-ATP Rs1 state (PDB: 4AAQ, chain A) to subunits 1, 3, 4 and 6. We then fitted the crystal structure of a GroES-bound GroEL subunit (derived from PDB: 1SVT) to the density of subunits 2, 5, and 7. We optimised the model using ISOLDE (Croll, 2018). Resolution for the equatorial domains allowed for confident modelling of the C_α backbone and some amino acid side chains. Intermediate and apical domains were less well resolved. In these regions we applied distance restraints to the model prior to optimisation. Density for non-native substrate was left unmodelled.

Both structures showed GroEL subunits 1, 3, 4 and 6 interacting with non-native substrate primarily via their lower hydrophobic helix (helix I) (**Fig. 6-11A** and **Fig. 6-11B**, left). The conformation of these four subunits did not deviate far from the Rs1 state of GroEL-ATP (Clare et al., 2012). The low local resolution of the GroEL apical domains limited our interpretation of specific residues contacting Rubisco. However, the interaction appeared broadly similar to that seen in our reconstructions of nucleotide-free GroEL-Rubisco (**Section 5.3.3** and **Section 5.4.4**). GroEL subunits 2, 5 and 7 had adopted a conformation similar to the GroES-bound state. As noted previously, this conformation has not been observed in the absence of GroES. This arrangement of GroEL subunits would allow four subunits to sequester non-native substrate while the other three simultaneously recruit GroES.

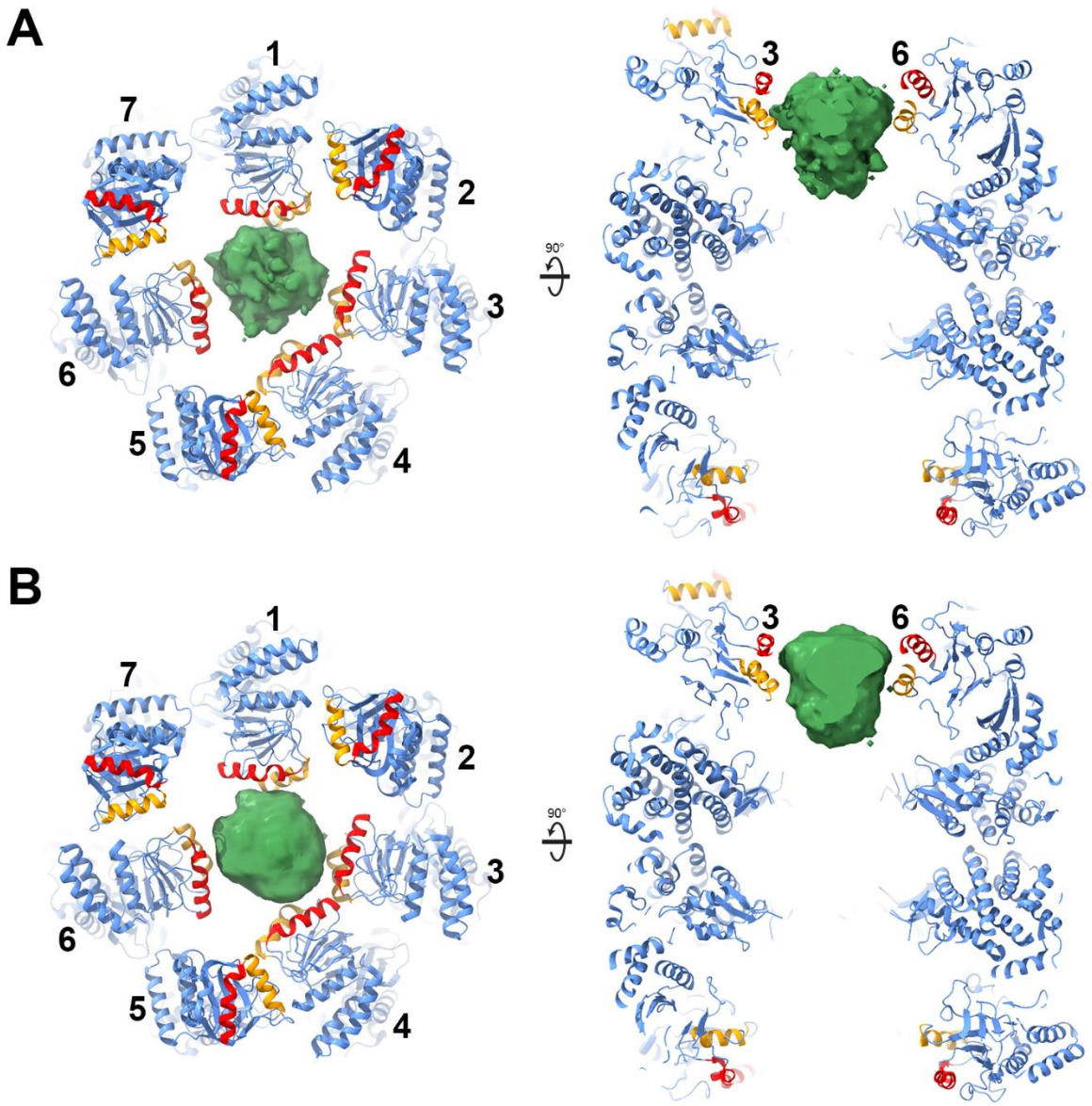

Figure 6-11. Molecular models of GroEL -ADP·BeF_x (top) and GroEL -ADP·AlF_x (bottom) with unmodelled Rubisco. GroEL is depicted as a ribbon diagram and coloured blue. Helices H and I are coloured red and orange.

6.2.6. Comparison with the crystal structure of GroEL-GroES

Of the three GroEL subunits in a GroES-bound conformation, subunit 7 was the best resolved for both reconstructions. We therefore compared our model of subunit 7 (**Fig. 6-12A**) to a GroEL subunit from the crystal structure of GroEL-GroES-ADP·AlF_3 (PDB: 1SVT) (**Fig. 6-12B**). The equatorial and intermediate domains matched well with the crystal structure, giving RMSD values

of 1.2 – 1.3 Å. The apical domains matched less well with an RMSD of 4.3 Å. A superposition of the two models showed that the apical domains in our structure of subunit 7 were in a lower, slightly less extended conformation (**Fig. 6-12C**). However, local resolution here was relatively low. A model derived from a larger cryoEM data set was needed to make a more valid comparison.

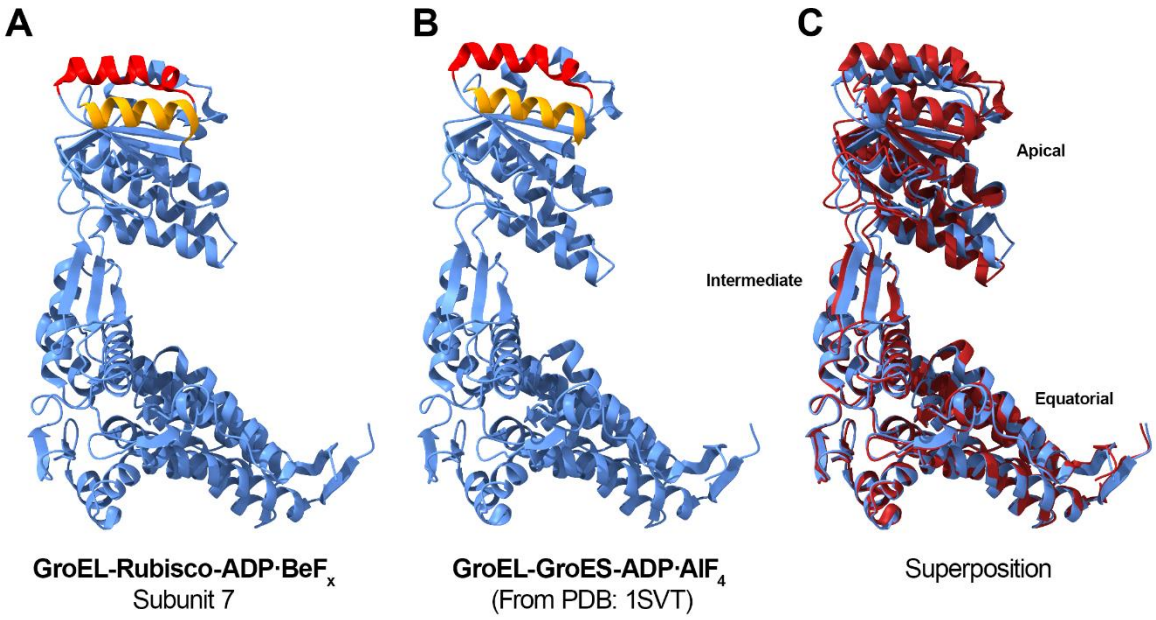

Figure 6-12. Comparison of GroEL subunits from GroEL-Rubisco-ADP·BeF_x and GroEL-GroES-ADP·AlF_x. (**A**) Isolated single subunit from the cryoEM structure of GroEL-Rubisco-ADP·BeF_x. (**B**) Single subunit from the crystal structure of GroEL-GroES-ADP·AlF_3. (**C**) Superposition of single subunits from (**A**) and (**B**). GroEL is depicted as a ribbon diagram and coloured blue. Helices H and I are coloured red and orange.

6.3. High-resolution structure of GroEL-Rubisco-ADP·BeF_x

We next aimed to determine a high-resolution structure of the complex to better resolve GroEL-substrate interactions, nucleotide states and other insights. We had not observed significant structural differences between our cryoEM reconstructions using either of the ATP analogues. Both analogues trapped the same asymmetric structure of GroEL and both data sets yielded reconstructions with visible density for substrate.

In the canonical GroEL-GroES ATPase cycle, ATP hydrolysis occurs after GroES has bound (see **Fig. 1-6** in chapter I). We were interested in studying the complex after ATP binding and prior to ATP hydrolysis. We therefore decided to focus on GroEL-Rubisco-ADP·BeF_x instead of the ADP·AlF_x complex. To minimise the interaction of GroEL with the air-water interface, we prepared grids for cryoEM using the Chameleon instrument (SPT Labtech) with assistance from Michele Darrow.

6.3.1. CryoEM of GroEL-Rubisco-ADP·BeF_x

Using a Chameleon instrument to prepare grids for cryoEM, we achieved dispense-to-freeze times of 50 – 100 ms. Based on our previous experience with the Chameleon instrument, we prepared the sample at a concentration of ~6 mg/mL. For our previous samples, a concentration in the range of 5 – 7 mg/mL had provided an ideal particle distribution. Micrographs typically contained > 300 particles each. However, cryoEM images of GroEL-Rubisco-ADP·BeF_x at ~6 mg/mL showed a far above-ideal particle concentration. We observed significant crowding and overlapping of particles. We therefore were forced to discard more particles than usual during our initial image processing steps.

6.3.2. Initial processing

Two cryoEM data sets of GroEL-Rubisco-ADP·BeF_x were collected from Chameleon grids prepared in the same session. Micrographs from both data sets showed crowded and overlapping particles (**Fig. 6-13A** and **Fig. 6-13B**). End views of GroEL seemed more prone to overlapping than side views. Initial 2D classification yielded multiple class averages showing overlapping end views (**Fig. 6-13C** and **Fig. 6-13D**, top row). After three rounds of 2D classification in cryoSPARC, it appeared we had removed most overlapping particles. Class averages showed clear secondary structure (**Fig. 6-13C** and **Fig. 6-13D**, bottom row). Particles were extracted at 3.2 Å/px and refined (**Fig. 6-13E** and **Fig. 6-13F**). The consensus map showed the asymmetric structure of

GroEL seen in previous GroEL-Rubisco-nucleotide data sets (**Fig. 6-13E** and **Fig. 6-13F**, top). Angular distribution plots clearly showed the preferred orientation of particles (**Fig. 6-13G** and **Fig. 6-13H**).

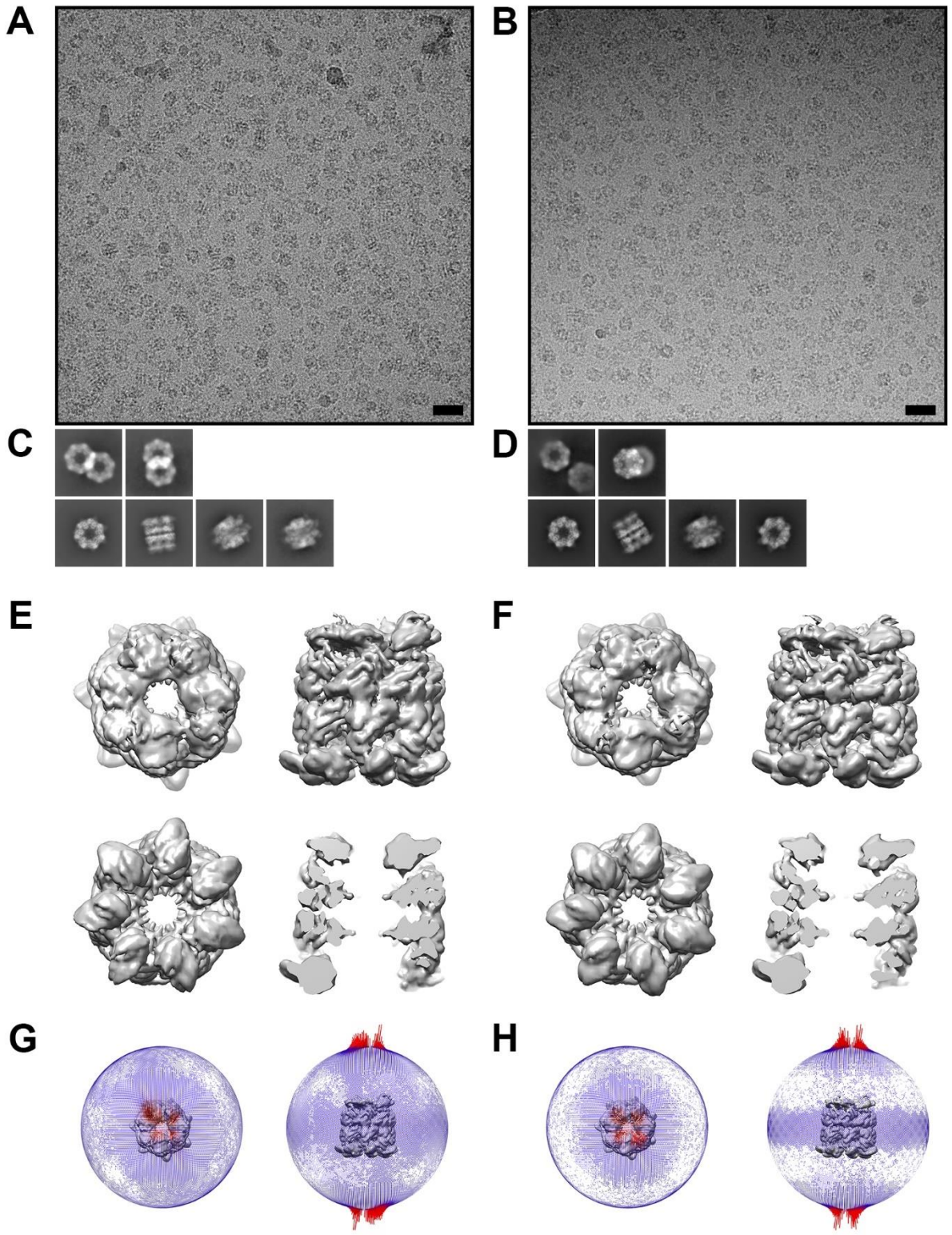

Figure 6-13. Consensus reconstructions from the two GroEL-Rubisco-ADP·BeF_x data sets. Representative motion-corrected cryo micrographs of data set 1 (**A**) and data set 2 (**B**) (top; black scale bar = 20 nm). 2D class averages showing overlapping particles, and a range of views of the GroEL-Rubisco-ADP·BeF_x from data set 1 (**C**) and data set 2 (**D**). Consensus reconstructions from data set 1 (**E**) and data set 2 (**F**). Viewed from the *cis* ring (top left), side (top right), *trans* ring (bottom left) and a central slice (bottom right). 3D histogram showing the orientation distribution of GroEL-Rubisco-ADP·BeF_x in data set 1 (**G**) and data set 2 (**H**).

6.3.3. Image processing

Image processing was performed in two main steps. First, we used 3D classification to identify and combine the best-resolved particles from each data set. We also removed remaining overlapping particles that were missed during 2D classification (**Fig. 6-14**). We then performed focussed 3D classification to identify substrate-bound particles (**Fig. 6-15**).

Particles from 2D classification were extracted at 3.2 Å/px and refined to 6.4 Å (**Fig. 6-14A** and **Fig. 6-14B**). Particles were subjected to CTF refinement to estimate beam tilt, anisotropic magnification, and per-particle defocus. We then performed 3D classification (**Fig. 6-14C** and **Fig. 6-14D)**. At this stage, we removed over a million poorly resolved particles from each data set. To better understand the features in these particles, we performed 2D classification on the discarded classes. The resulting 2D class averages suggested that almost all of the discarded particles were overlapping end views of GroEL (not shown). Particles from the best 3D class were then re-extracted at 1.95 Å/px and refined to 4.4 – 4.5 Å (**Fig. 6-14E** and **Fig. 6-14F**). CTF refinement and further 3D refinement improved the resolution to 3.9 Å for both data sets (**Fig. 6-14G** and **Fig. 6-14H**). Particles were subjected to another round of 3D classification using a soft-edge solvent mask encompassing the whole complex (**Fig. 6-14I** and **Fig. 6-14J**). Poorly resolved classes were discarded and particles from good 3D classes from each data set were

combined into a single class (**Fig. 6-14E** and **Fig. 6-14F**, green dashed box indicates good classes). The combined class contained 678k particles, 327k from data set 1 and 351k from data set 2.

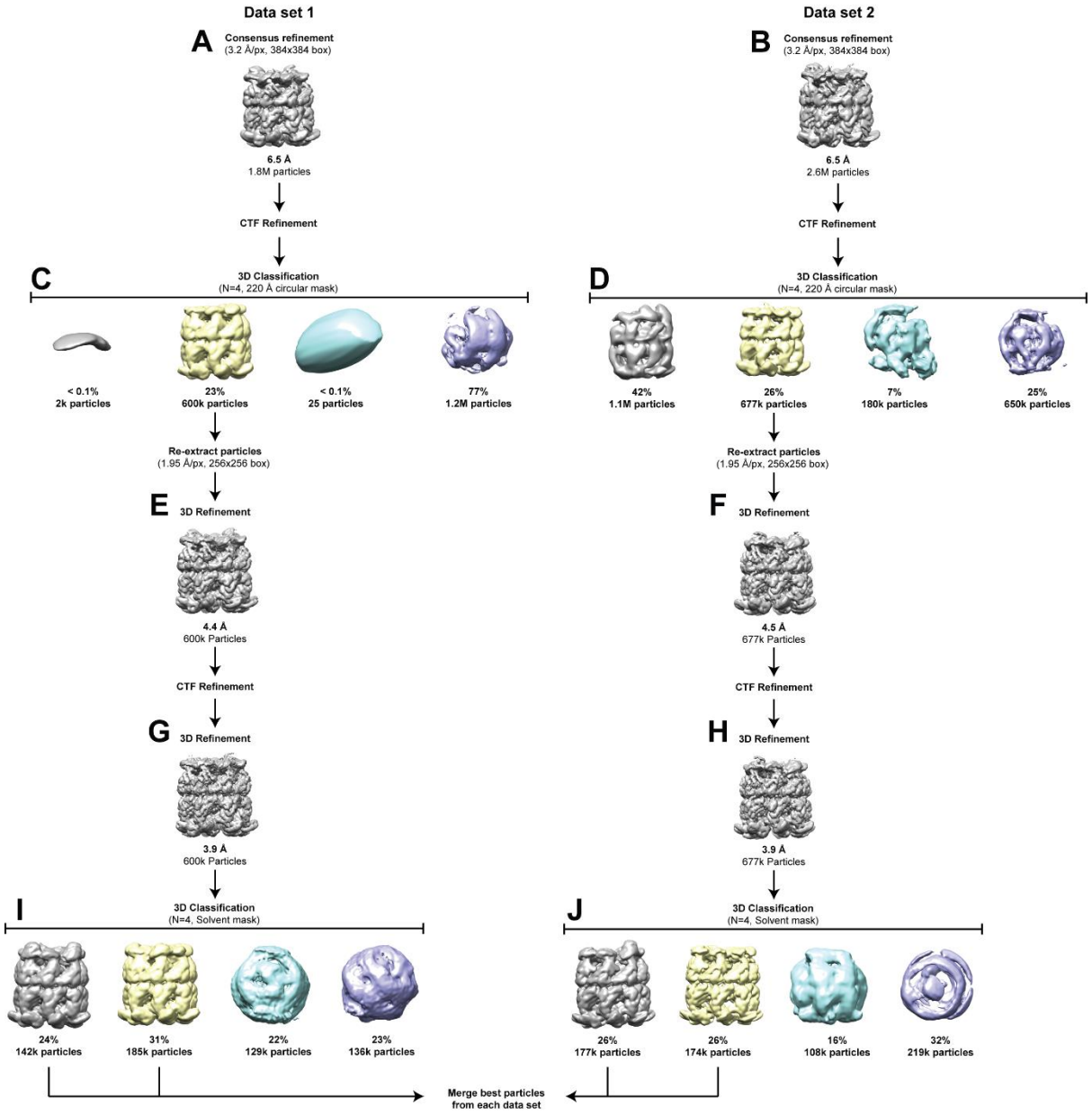

Figure 6-14. Initial 3D classification and merging of GroEL-Rubisco-ADP·BeF_x data sets. (**A** and **B**) Consensus reconstructions from each data set. (**C** and **D**) Initial 3D classification of particles in each data set, using default Relion settings and four classes. (**E** and **F**) Reconstructions from the best resolved 3D class. (**G** and **H**) Reconstructions after CTF refinement. (**I** and **J**) 3D classification using a solvent mask around the entire GroEL complex. The best classes were merged and taken forward for further processing.

6.3.4. Focussed 3D classification

We used focussed 3D classification to identify particles with strong substrate density (**Fig. 6-15**). The final merged class of particles from **Fig. 6-14** was refined, yielding a 3.9 Å map (**Fig. 6-15A**). At this stage, we extracted all particles with a smaller pixel size, then performed CTF refinement, particle polishing and 3D refinement. The resulting 2.9 Å map was used for model building of the GroEL equatorial and intermediate domains (**Fig. 6-15B**). We split the down-sampled (1.95 Å/px) particles into two subsets to speed up computation (**Fig. 6-15C**). Each subset of 339k particles was subjected to focussed 3D classification using a mask that encompassed the GroEL cavity (**Fig. 6-15D** and **Fig. 6-15E**). Approximately half of the particles in each subset showed no substrate density (**Fig. 6-15D**, orange, class 4; **Fig. 6-15E**, yellow, class 2). Classes that showed strong substrate density were combined into a single class and refined to 3.9 Å (**Fig. 6-15F**). We also combined particles from both apo classes and refined them to 3.9 Å (**Fig. 6-15G**). Substrate-bound particles showed unresolved heterogeneity for two of the GroEL subunits. At low threshold values, subunits 4 and 5 appeared to be in substrate-bound and GroES-bound conformations simultaneously. To resolve this, we created a mask around the apical domains of subunits 4 and 5 (**Fig. 6-15H**). We used this mask in focussed 3D classification without alignments (**Fig. 6-15I**). Two classes of ~100k particles each suggested alternate states for subunits 4 and 5 (**Fig. 6-15I**, grey and blue classes). We re-extracted particles from classes 1 and 4 at a higher sampling frequency. We then performed CTF refinement, particle polishing and 3D refinement on each class. Both classes yielded 3.5 Å reconstructions (**Fig. 6-15J** and **Fig. 6-15K**). Reconstructions from classes 2 and 3 had poorly resolved density for subunits 4 and 5 (not shown). The reconstructions from classes 1 and 4 showed essentially the same conformation of the complex, related by an anticlockwise rotation through two GroEL subunits ($360°/7 * 2 = 103°$) about the 7-fold symmetry axis (illustrated in **Fig. 6-15L**). We merged classes 1 and 4 and refined the particles to a final resolution of 3.4 Å (**Fig. 6-15M**).

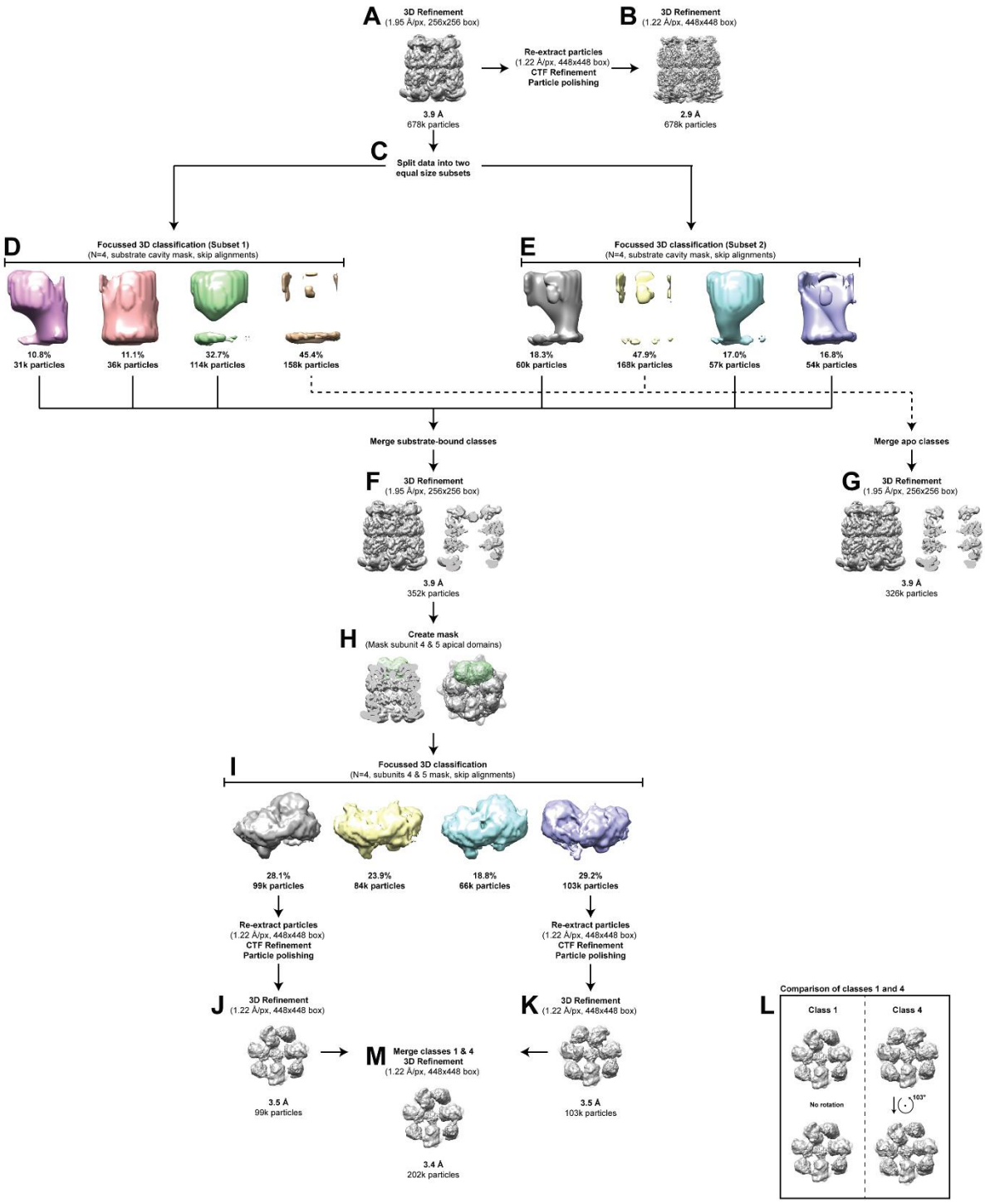

Figure 6-15. Focussed 3D classification of GroEL-Rubisco-ADP·BeF$_x$. (**A**) Reconstruction from the best particles identified during initial image processing. (**B**) High resolution refinement after re-extraction of particles, CTF refinement and Bayesian polishing. (**D** and **E**) Focussed 3D classification of substrate density. (**F**) Reconstruction of substrate-bound particles. (**G**) Reconstruction of substrate-free particles. (**H**) Creation of a mask to encompass apical domains

of GroEL subunits 4 and 5. (**I**) Focussed 3D classification of GroEL subunit 4 and 5 apical domains. (**J** and **K**) Reconstructions from the best two classes, related by the rotation depicted in (**L**). (**M**) Final reconstruction of GroEL-Rubisco-ADP·BeF_x.

6.3.5. CryoEM reconstruction of GroEL-Rubisco-ADP·BeF_x

Our high resolution cryoEM reconstruction remained consistent with our earlier reconstructions of GroEL-Rubisco-ATP and GroEL-Rubisco-ADP·BeF_x/AlF_x. We observed the same asymmetric arrangement of *cis* ring subunits, with four GroEL subunits contacting non-native Rubisco (**Fig. 6-16A**, top left). As before, three GroEL subunits adopted a conformation similar to the GroES-bound state and were only fully resolved at a low density threshold (**Fig. 6-16A**, top right, grey silhouette and black arrowheads). The GroEL *trans* ring showed approximate 7-fold symmetry (**Fig. 6-16A**, bottom left). Interactions between GroEL and non-native Rubisco were again mainly located in helix I. Non-native Rubisco also contacted the C-terminal tails of all seven GroEL subunits (**Fig. 6-16A**, bottom right). Local resolution was highest in the equatorial and intermediate domains, with many regions showing clear density for amino acid sidechains (**Fig. 6-16B**). For subunits in the GroES-bound conformations, density for helices H and I was only resolved at low threshold values. The global resolution was estimated at 3.4 Å (**Fig. 6-16C**). Angular distribution plots showed that most particles had adopted end and side orientations (**Fig. 6-16D**).

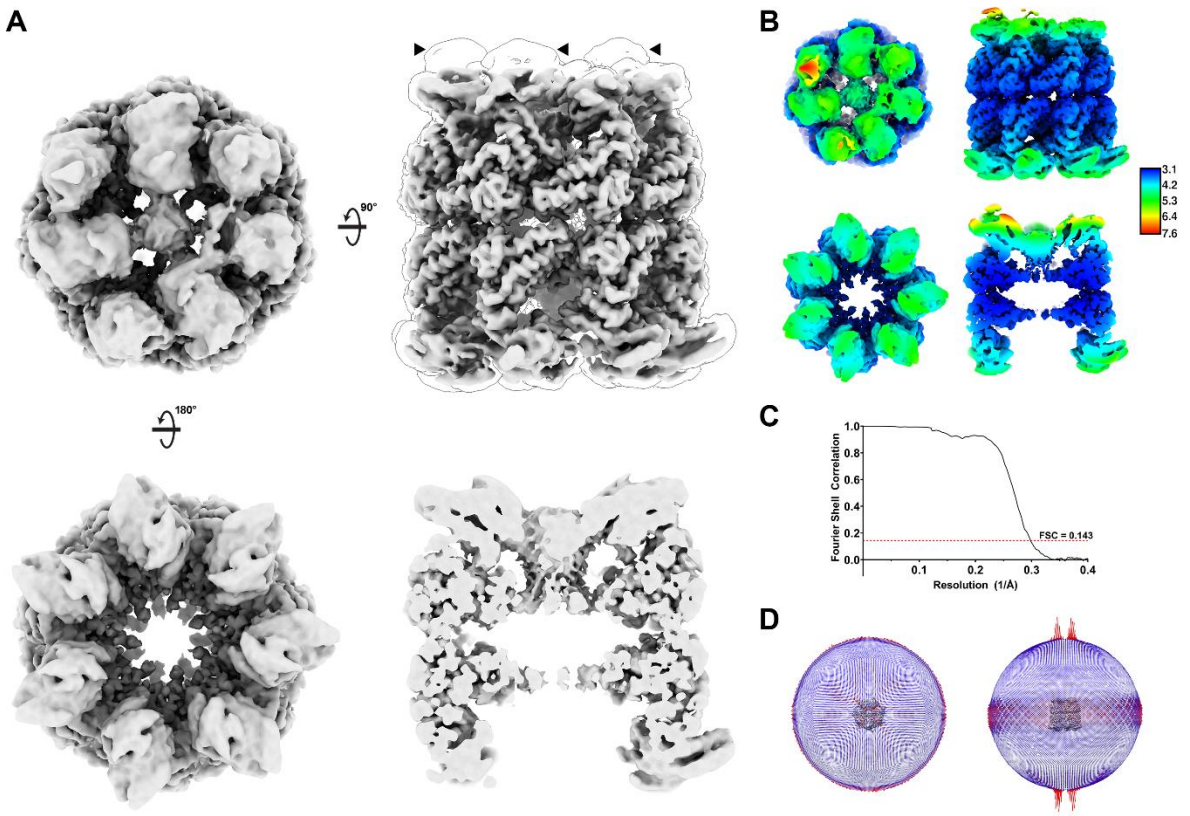

Figure 6-16. Reconstruction of GroEL-Rubisco-ADP·BeF_x shows non-native Rubisco contacting apical domains and C-termini of GroEL subunits. (**A**) 3D reconstruction of substrate-bound GroEL-ADP·BeF_x. In the side view depiction, a low threshold outline is shown to highlight the weaker density of GroEL subunit 2, 5 and 7 elevated apical domains. (**B**) Local resolution of GroEL-Rubisco-ADP·BeF_x. (**C**) Fourier shell correlation (FSC) plot for the calculated half-maps; overall resolution was 3.4 Å. (**D**) 3D histogram showing the orientation distribution of GroEL-Rubisco-ADP·BeF_x particles.

6.3.6. High resolution model building

We took advantage of previously published crystal structures of apo GroEL and GroEL-GroES during model building. To determine a molecular model of the *cis* ring, we rigid-body fitted the crystal structure of apo GroEL (PDB: 1SS8, chain A) to subunits 1, 3, 4 and 6, then optimised the model using ISOLDE (Croll, 2018). As before, we used distance restraints when modelling GroEL

apical domains as local resolution was lower. For subunits 2, 5 and 7, we performed rigid body fitting of GroEL subunits derived from the published crystal structure of GroEL-GroES-ADP·AlF_3 (PDB: 1SVT). The *trans* ring exhibited approximate 7-fold symmetry, we therefore modelled one subunit first. We fitted the crystal structure of apo GroEL (PDB: 1SS8, chain A) to the density of a single *trans* ring subunit. We then optimised the fit using ISOLDE and copied the optimised model to the other six *trans* ring subunits. The whole model was then optimised further in ISOLDE to improve the model-to-map fit. The final model is shown in **Fig. 6-17**. Selected areas of the map and model are shown in **Fig. 6-18**.

6.3.7. Estimation of non-native Rubisco mass

After model building, we measured the volume of non-native Rubisco as before and estimated its mass to be 24.5 kDa (see **Section 5.4.6** for mass estimation method). This represented approximately 50% of a Rubisco monomer (50.5 kDa). We observed no additional substrate density protruding from the GroEL *cis* cavity.

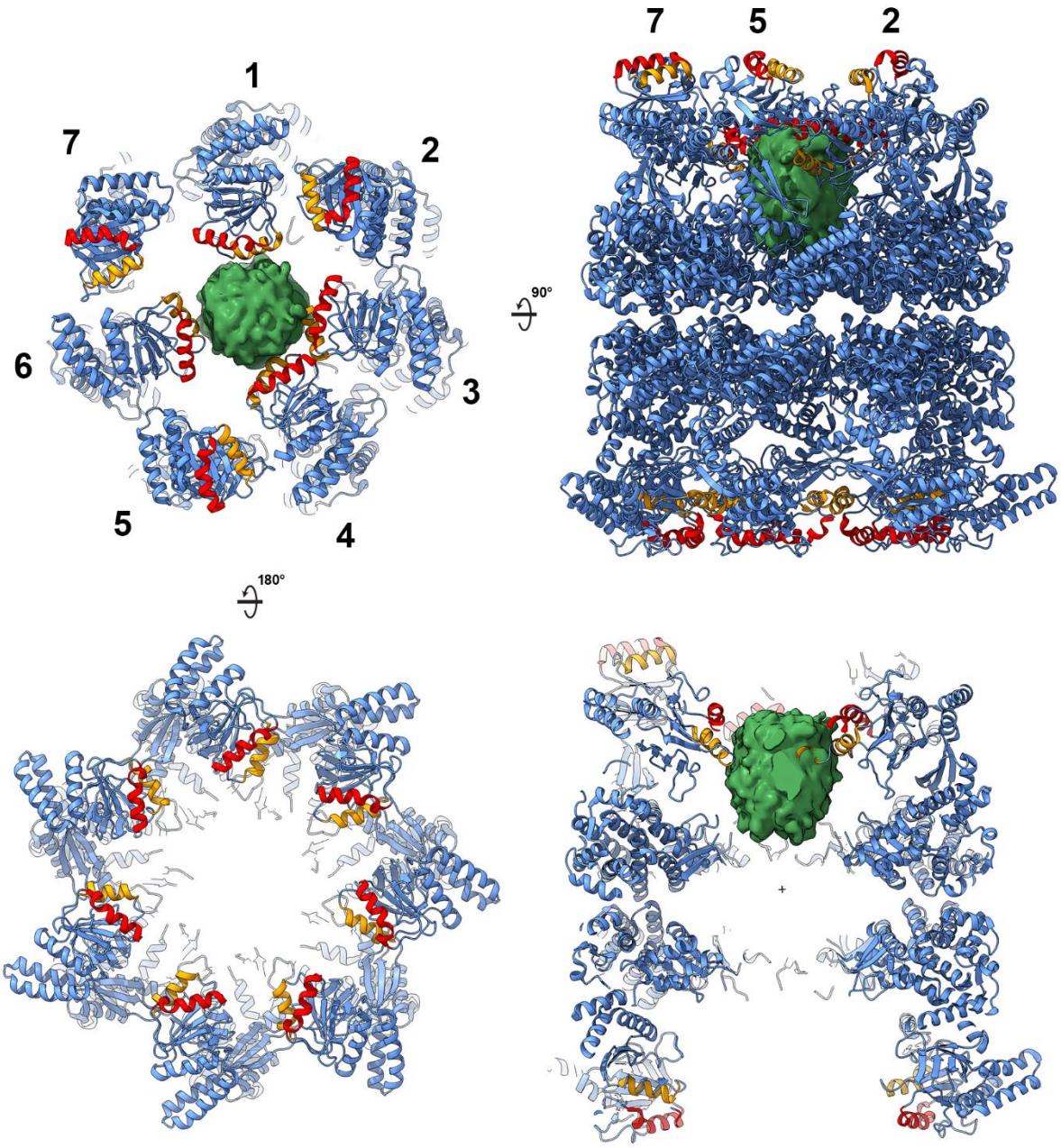

Figure 6-17. CryoEM structure of GroEL-Rubisco-ADP·BeF$_x$. GroEL is depicted as a blue ribbon diagram, with helices H and I coloured red and orange. CryoEM density for bound non-native Rubisco is shown in green.

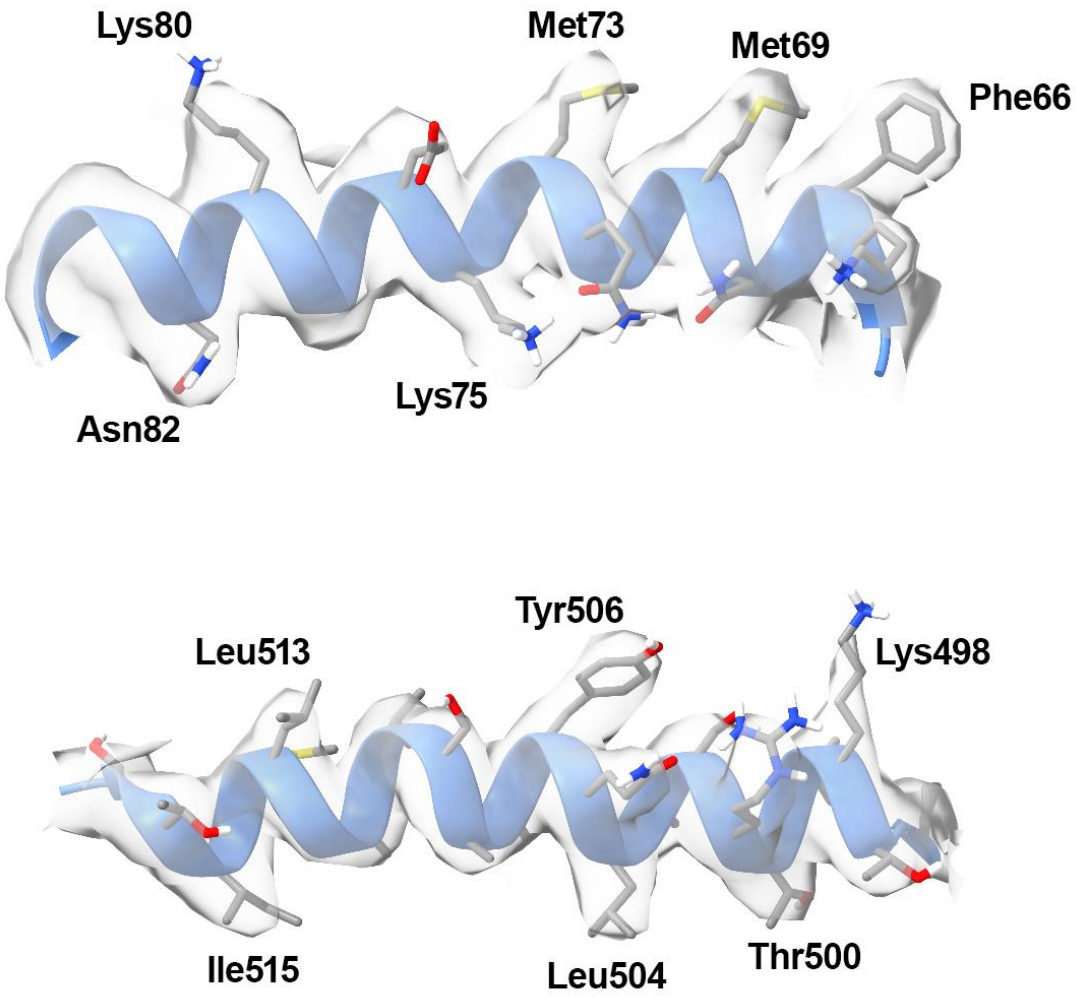

Figure 6-18. CryoEM map quality. Isolated density of two GroEL helices illustrating the map-model fit.

6.3.8. Nucleotide states of *cis* and *trans* ring subunits

We examined the ATP binding sites of GroEL subunits (**Fig. 6-19**). We observed differences in nucleotide density between *cis* and *trans* rings but saw no significant differences between subunits of the same ring. Strong density for ADP was seen in all fourteen sites. We observed continuous density between the D87 sidechain and the ADP β-phosphate. The residue D87 is involved in ATP hydrolysis, and mutations such as D87K abolish ATPase activity (Fenton et al., 1994). At a local resolution of 3.1 Å, we were able to confidently model ADP and the phosphate oxygen-coordinating metal, Mg^{2+}. Nucleotide bound in the *cis* ring showed additional density that we

attributed to the ATP γ-phosphate analogue BeF_3 (**Fig. 6-19A**). Nucleotides bound in the *trans* ring lacked this additional density and were modelled as ADP without BeF_3 (**Fig. 6-19B**). The overall structure is therefore GroEL-Rubisco-$(ADP{\cdot}BeF_3)_7$-ADP_7. *Cis* ring subunits showed the presence of the second coordinating metal, K^+ (**Fig. 6-19A**, purple sphere). GroEL requires K^+ to hydrolyse ATP and a published crystal structure has confirmed this position as the K^+ binding site (Kiser et al., 2009). In the *cis* ring, we also observed extra density between the D52 and D398 sidechains (**Fig. 6-19A**). This presumably represented the water molecule involved in attacking the γ-phosphate of ATP during hydrolysis. This is based on the mechanism of ATP hydrolysis deduced from the structure of the *T. acidophilum* thermosome bound to $ADP{\cdot}AlF_3$ (Ditzel et al., 1998).

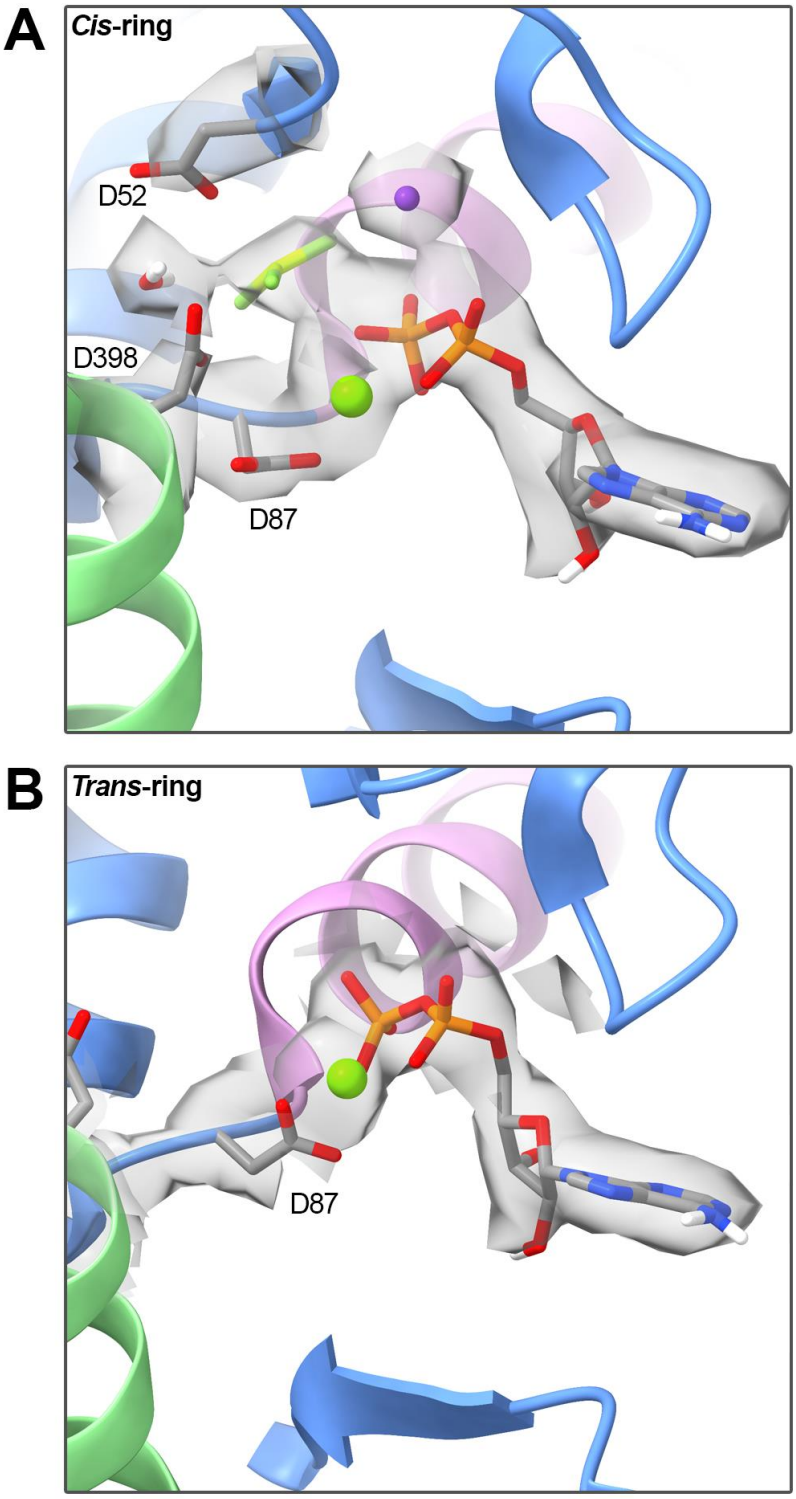

Figure 6-19. The nucleotide binding site of the *cis* and *trans* rings of GroEL-Rubisco-ADP·BeF$_x$. (**A**) View of the *cis* ring nucleotide binding site, with bound ADP, Mg^{2+} (green sphere), BeF_x (green stick), K^+ (purple sphere) and catalytic water. (B) View of the *trans* ring nucleotide binding site, showing bound ADP-Mg.

6.3.9. Mapping interactions between GroEL-ADP·BeF_x and non-native Rubisco

Non-native Rubisco interacted with the apical domains of four GroEL subunits, two consecutive and two non-consecutive (**Fig. 6-20A**). Compared to the nucleotide-free binary complex, Rubisco showed stronger contacts to the disordered GroEL C-termini (**Fig. 6-20A**, middle and right). The strongest interactions between GroEL-ADP·BeF_x and Rubisco were similar to those observed in the binary complex. The same Arg-268 contact at the C-terminus of helix I was seen for all four interaction sites (**Fig. 6-20B-E**). Other subunits showed contacts via Val-264 (**Fig. 6-20B** and **Fig. 6-20E**). We observed typically weaker interactions with Tyr-203 of the underlying hydrophobic segment (**Fig. 6-20E**). Other contacts were apparent at lower density thresholds, but due to the low local resolution, we were unable to confidently identify specific residues. Most cavity-facing residues appeared to be involved in contacting non-native substrate at low thresholds.

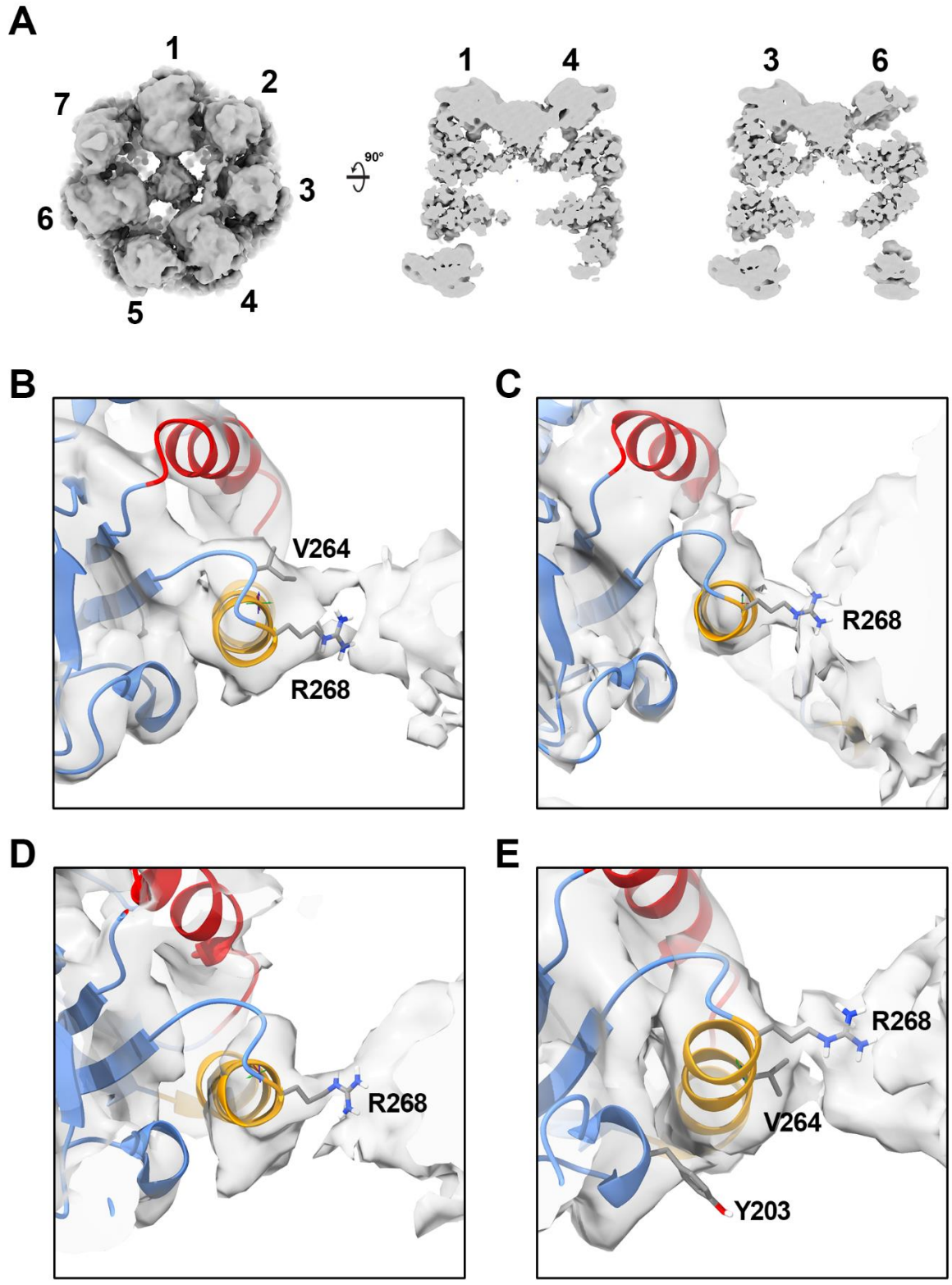

Figure 6-20. Interactions between non-native Rubisco and GroEL-ADP·BeF$_x$. (**A**) CryoEM map showing numbering of GroEL subunits. Interactions of non-native Rubisco are highlighted with subunit 1 (**B**), subunit 3 (**C**), subunit 4 (**D**) and subunit 6 (**E**).

6.3.10. Identification of an equatorial-apical salt bridge in substrate-bound GroEL subunits

The four GroEL subunits that contacted non-native Rubisco appeared to share an equatorial-to-apical domain salt bridge, R58-E209 (**Fig. 6-21A** and **Fig. 6-21B**). R58 is located within a short α-helix adjacent to the stem loop. E209 lies in a short loop region of the underlying segment. GroEL subunits in the GroES-bound conformation lack this salt bridge, as the distance between R58 and E209 is ~40 Å. We inspected our cryoEM structure of GroEL-Rubisco for this salt bridge. In nucleotide-free GroEL-Rubisco, E209 faces away from the R58 helix and their sidechains are ~8 Å apart. The loop region also adopts a different conformation (**Fig. 6-21C**). This conformation is also seen in the crystal structure of apo GroEL (PDB: 1SS8).

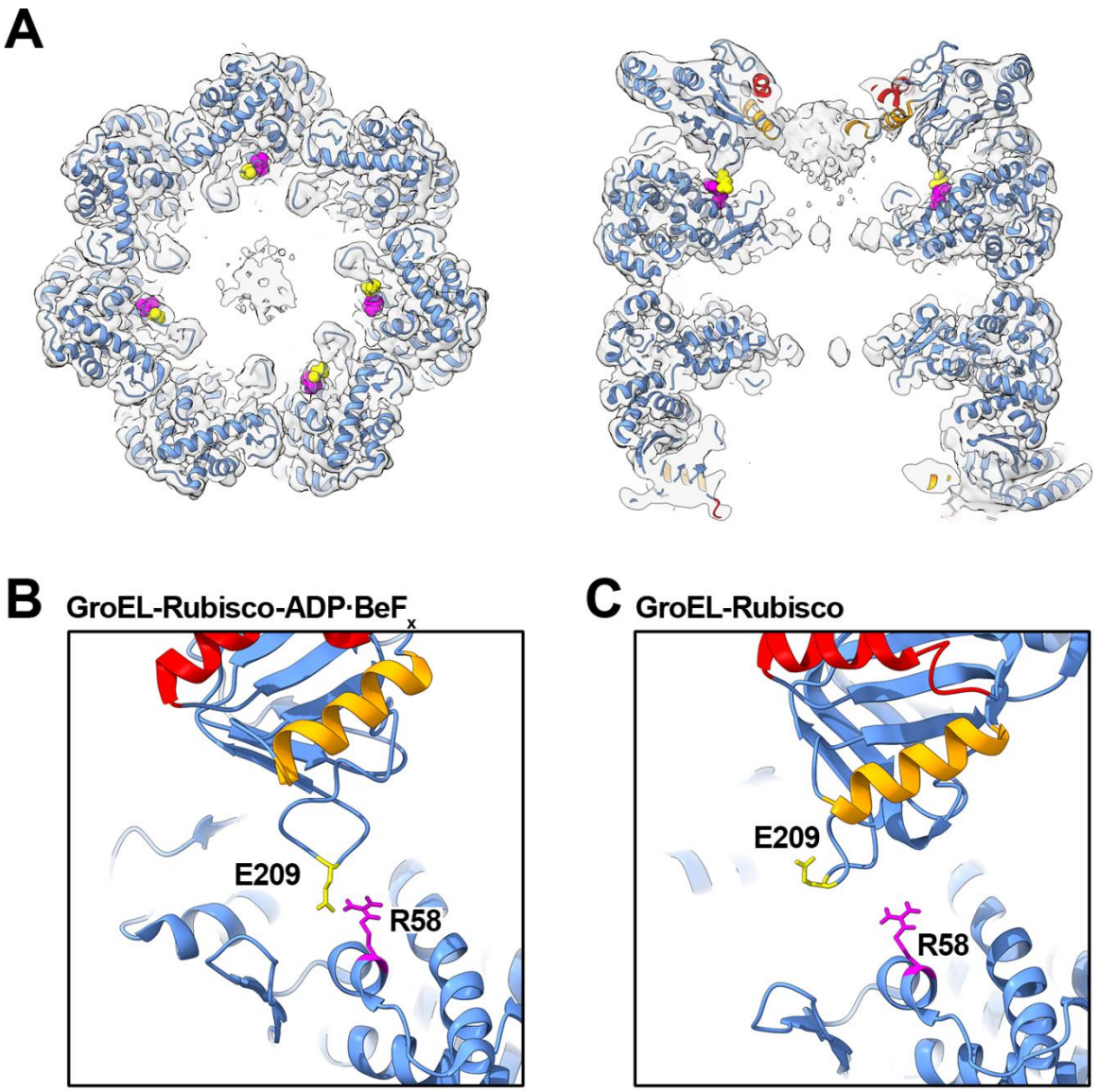

Figure 6-21. Salt bridge R58-E209. (**A**) Slice through the *cis* ring (left) and side view (right) highlighting the plane and location of the putative salt bridge. Helices H (red) and I (orange) are coloured. (**B**) Model of the R58-E209 salt bridge in GroEL-Rubisco-ADP·BeF_x. (**C**) R58-E209 residues in our GroEL-Rubisco cryoEM model, showing the E209 loop facing away. R58 and E209 atoms are coloured magenta and yellow respectively.

Chapter VII

7. CryoEM study of GroEL-GroES-Rubisco

Our cryoEM structure of GroEL-Rubisco-ADP·BeF_x revealed a novel asymmetric arrangement of GroEL apical domains. This conformation of GroEL can recruit the co-chaperonin GroES while maintaining contact with bound substrate. We aimed to study substrate encapsulation and folding upon binding of GroES. We formed binary complexes of GroEL bound to non-native Rubisco, then added ADP·AlF_x and GroES to form stalled folding-active ternary complexes.

GroEL-GroES complexes have been studied by cryoEM, both with and without substrate. The 10 Å reconstruction of GroEL-gp31-gp23 showed folded gp23 within the *cis* cavity of GroEL-gp31.ADP·AlF_3 (gp31 is a GroES homologue from bacteriophage T4). The reconstruction also showed non-native gp23 bound to the *trans* ring (Clare et al., 2009). Additionally, a 9.2 Å reconstruction of GroEL43Py398A-GroES-Rubisco-ATP showed density for non-native Rubisco in the *cis* cavity (Chen et al., 2013). However, GroEL43Py398A-GroES is not a fully folding-active complex. It encapsulates non-native substrate but is very slow to release it into the *cis* cavity. This reconstruction showed non-native Rubisco interacting with the cavity walls of GroEL. Specifically, the authors observed interactions with the equatorial and apical domains of GroEL. However, the interaction between GroEL and non-native substrate might have been altered by the pyrene dye attached covalently to residue C43.

We collected a high resolution cryoEM data set of GroEL-GroES-Rubisco from grids prepared using a Chameleon instrument. This dataset showed GroES-free and GroES-bound reconstructions of GroEL-Rubisco-ADP·AlF_x. GroES-free complexes were similar to our cryoEM structure of GroEL-Rubisco-ADP·BeF_x. GroES-bound complexes showed an

encapsulated Rubisco monomer in a near-native conformation interacting with the inner cavity walls of GroEL-GroES (**Section 7.1**).

7.1. GroEL-GroES-Rubisco-ADP·AlFx

We prepared cryoEM grids of GroEL-GroES-Rubisco-ADP·AlF_x using a Chameleon with the assistance of Michele Darrow. During grid preparation, GroES and ADP·AlF_x were added to binary complexes around ten minutes prior to loading the sample into the Chameleon. Grids were prepared during the same session as GroEL-Rubisco-ADP·BeF_x and we unfortunately observed the same crowding and overlapping of particles (see **Section 6.3.1**).

7.1.1. Initial processing

Motion-corrected micrographs showed crowded and overlapping particles (**Fig. 7-1A**, top). Initial 2D classification yielded class averages showing $GroEL_{14}$-$GroES_1$ "bullet" complexes (**Fig. 7-1A**, bottom). After three rounds of 2D classification most overlapping particles had been removed (although as with our previous data set, we removed more during subsequent 3D classification). We refined all particles from good 2D classes to calculate a consensus map. The consensus map showed well-resolved density for GroEL-GroES (**Fig. 7-1B**). Angular distribution plots showed the preferred orientation of end-views (**Fig. 7-1C**). The global resolution was Nyquist-limited at 6.8 Å (The FSC plot is omitted because at a pixel size of 3.4 Å, the FSC between the two half-maps reached Nyquist frequency before dropping below 0.95).

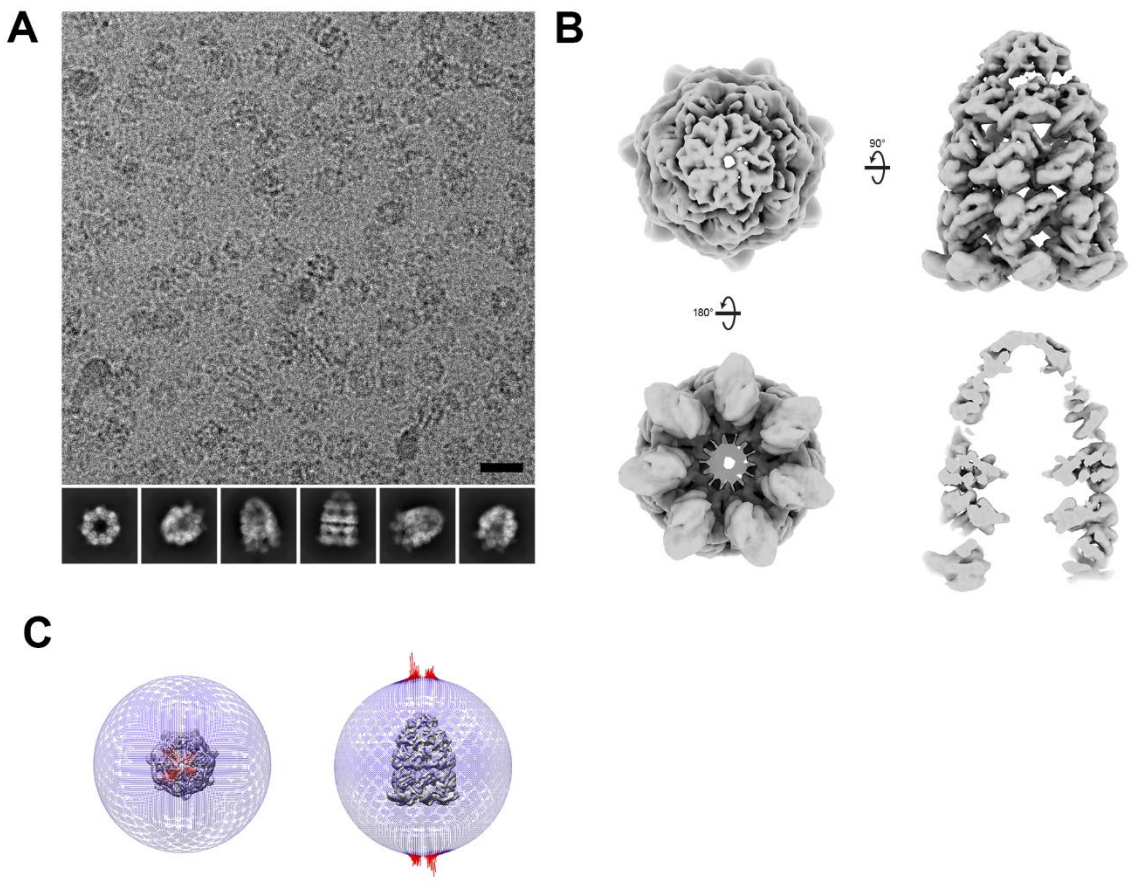

Figure 7-1. Consensus reconstruction of GroEL-GroES-Rubisco-ADP·AlF$_x$. (**A**) Representative micrograph area and 2D class averages. (**B**) Consensus reconstruction from 3.6M particles. (**C**) Orientation distribution plot generated by Relion and viewed in UCSF Chimera (Pettersen et al., 2004).

7.1.2. Further processing

Image processing of GroEL-GroES-Rubisco is divided into two parts. We first used rounds of unmasked 3D classification to identify the best resolved GroEL-GroES particles (**Fig. 7-2**). We then used masked 3D classification to identify particles with strong substrate density (**Fig. 7-3**). We initially tried to process the data using 3D classification with alignments, but all particles converged into a single class. We instead used 3D classification without alignments. We split the data into two subsets to speed up calculations. We saw several different 3D classes. One class in each subset showed strong density for bound GroES (**Fig. 7-2D**, class 1 and **Fig. 7-2E**, class 4).

Three classes in each subset showed the GroES-bound conformation of GroEL but showed weaker density for GroES. These classes hinted at incomplete GroES binding (**Fig. 7-2D**, classes 2, 4 and 5 and **Fig. 7-2E**, classes 2, 7 and 8). Additionally, we saw GroES-free classes that appeared similar to our reconstructions of GroEL-Rubisco-ADP·BeF_x/AlF_x (**Fig. 7-2D**, class 6 and **Fig. 7-2E**, class 3). This was likely because we had unintentionally used a sub-stoichiometric amount of GroES relative to GroEL. This was probably caused by an underestimation of the true GroEL concentration when preparing GroEL-GroES complexes for cryoEM study. A previous cryoEM data set of GroEL-GroES-Rubisco using a large excess of GroES did not produce any GroES-free classes (data not shown). Several poorly resolved low-population classes were identified and discarded. We experimented with different approaches of merging classes for 3D refinement. Classes that showed partial density for GroES yielded anisotropic reconstructions and were not interpretable. 2D classification suggested that these classes consisted almost entirely of end-view particles (not shown). Instead, we took a conservative approach and merged all classes that showed partial or full density for GroES, then subjected the combined class to 3D classification with alignments (**Fig. 7-2F**). One class containing a quarter of the particles showed a well-resolved GroEL-GroES complex (**Fig. 7-2F**, class 1). Class 2 contained the remaining three quarters of particles. We were unable to calculate a good 3D reconstruction from class 2 particles. Again, 2D classification suggested that class 2 consisted almost entirely of end-view and overlapping particles (not shown). We re-extracted the 814k particles from class 1 at a smaller pixel size and refined them to 4 Å.

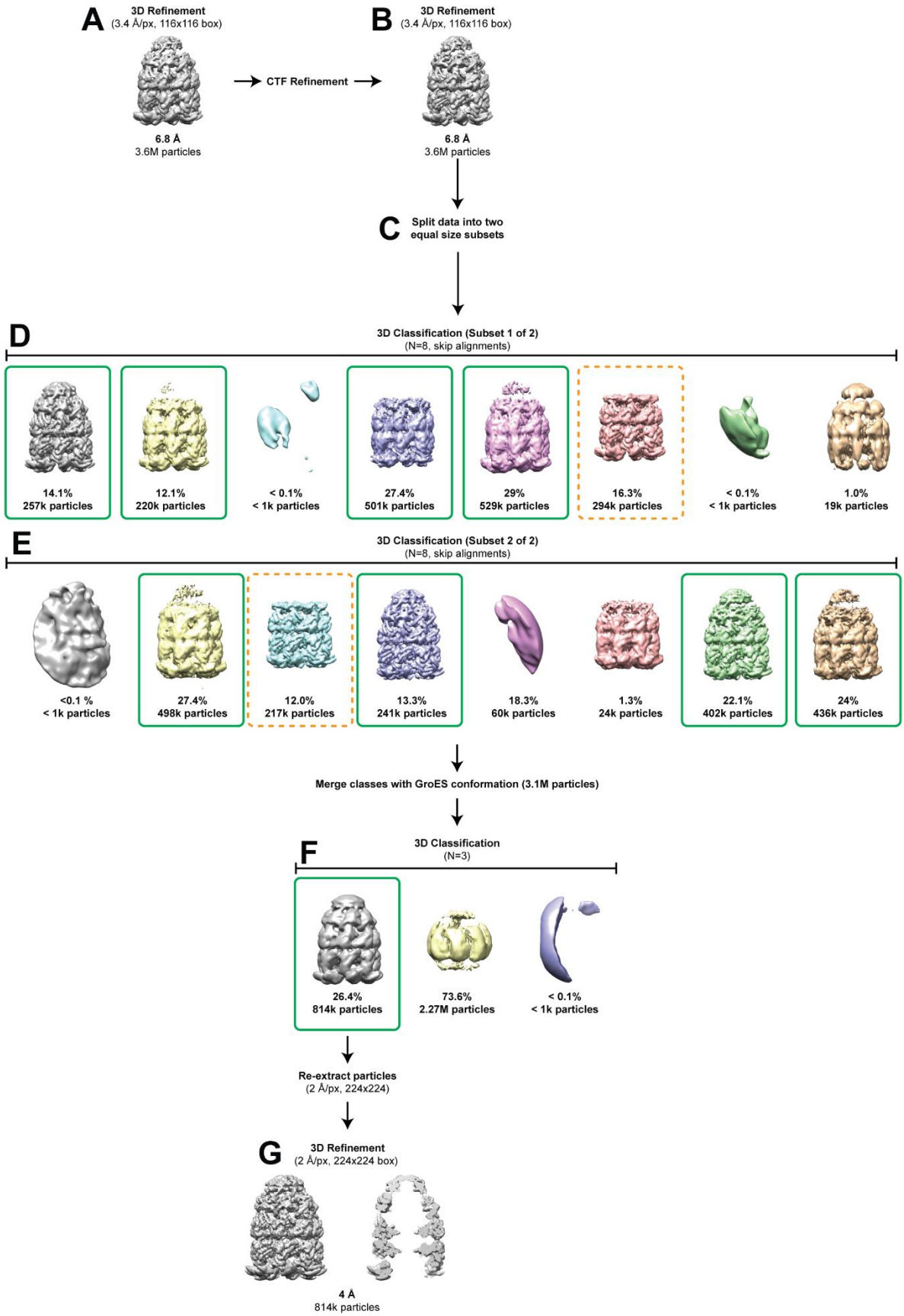
A
3D Refinement
(3.4 Å/px, 116x116 box)
6.8 Å
3.6M particles
CTF Refinement
B
3D Refinement
(3.4 Å/px, 116x116 box)
6.8 Å
3.6M particles
C
Split data into two
equal size subsets
D
3D Classification (Subset 1 of 2)
(N=8, skip alignments)
14.1%
257k particles
12.1%
220k particles
< 0.1%
< 1k particles
27.4%
501k particles
29%
529k particles
16.3%
294k particles
< 0.1%
< 1k particles
1.0%
19k particles
E
3D Classification (Subset 2 of 2)
(N=8, skip alignments)
<0.1 %
< 1k particles
27.4%
498k particles
12.0%
217k particles
13.3%
241k particles
18.3%
60k particles
1.3%
24k particles
22.1%
402k particles
24%
436k particles
Merge classes with GroES conformation (3.1M particles)
F
3D Classification
(N=3)
26.4%
814k particles
73.6%
2.27M particles
< 0.1%
< 1k particles
Re-extract particles
(2 Å/px, 224x224)
G
3D Refinement
(2 Å/px, 224x224 box)
4 Å
814k particles

Figure 7-2. Image processing of GroEL-GroES-Rubisco-ADP·AlF$_x$. (**A**) Consensus reconstruction from 3.6M particles. (**B**) Consensus reconstruction following CTF refinement. (**C**) Splitting of dataset into two 1.8M particle subsets. (**D**) 3D Classification of subset 1. (**E**) 3D Classification of subset 2. (**F**) 3D classification of all GroEL particles with partial or full density for GroES. (**G**) Reconstruction of the best resolved GroEL-GroES particles.

7.1.3. Substrate classification

We used the best resolved 814k GroEL-GroES particles for substrate classification (**Fig. 7-3**). To speed up calculations we again split the particles into two subsets, allowing them to be read into 128 GB of server RAM instead of reading from the slower disk drives (**Fig. 7-3A**). For each subset we used focussed 3D classification targeting the *cis* cavity of GroEL-GroES (**Fig. 7-3B** and **Fig. 7-3C**). Subsets totalling 31k particles showed strong density for bound substrate (**Fig. 7-3B**, class 4 and **Fig. 7-3C**, class 4). These subsets were merged and re-extracted at a smaller pixel size. Refinement of the re-extracted particles yielded a 4 Å reconstruction of substrate-bound GroEL-GroES (**Fig. 7-3D**). This improved to 3.8 Å following CTF refinement and further 3D refinement (**Fig. 7-3E**).

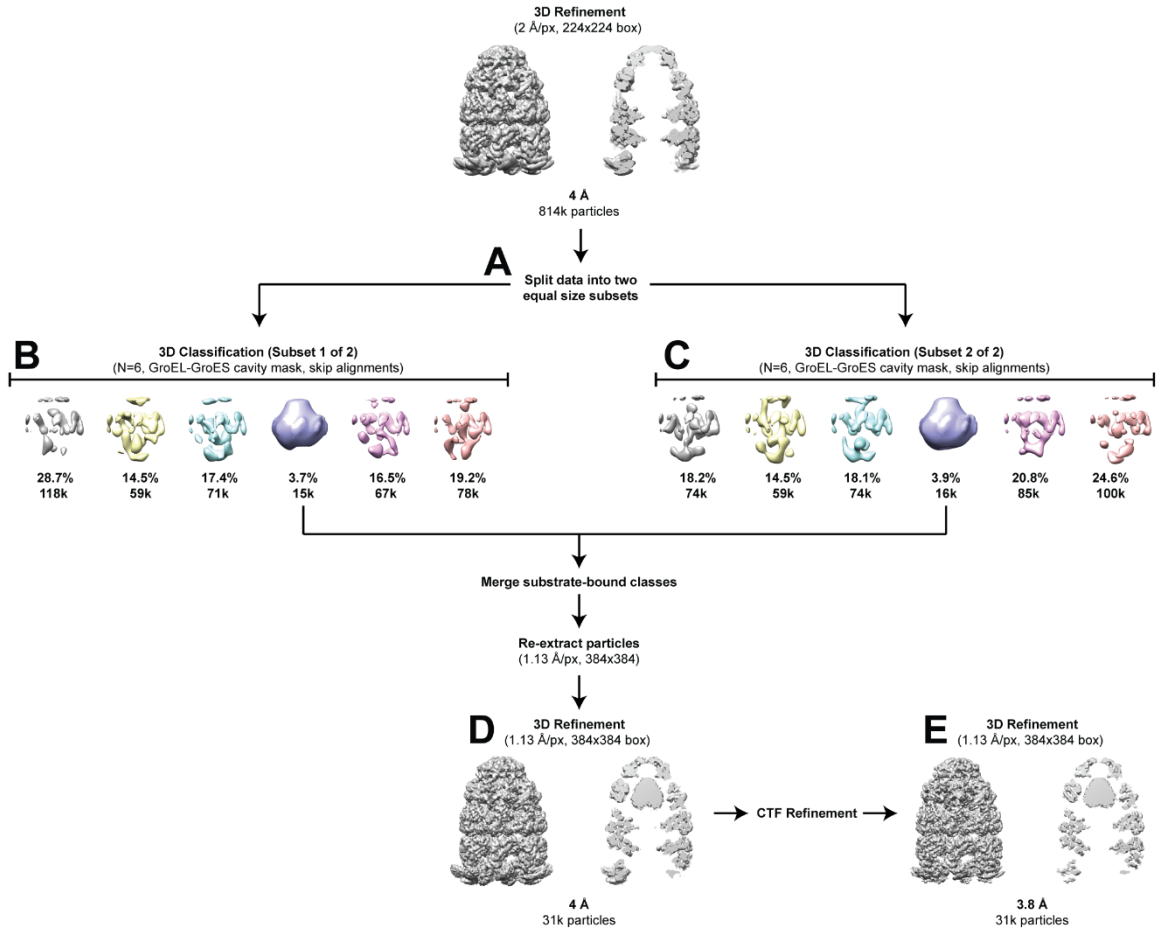

Figure 7-3. Further image processing of GroEL-GroES-Rubisco-ADP·AlF$_x$. (**A**) Splitting of the best resolved GroEL-GroES particles into two ~400k particle subsets. (**B**) Focussed 3D classification of subset 1. (**C**) Focussed 3D classification of subset 2. (**D**) Reconstruction of the substrate-bound GroEL-GroES particles. (**E**) Substrate-bound reconstruction following CTF refinement.

7.1.4. Substrate-bound reconstruction

The cryoEM reconstruction showed well-resolved density for GroEL-GroES that was similar to previously published crystal structures (**Fig. 7-4A**). Encapsulated Rubisco occupied around two-thirds of the *cis* cavity and appeared to interact with the GroEL-GroES cavity walls at discrete positions (**Fig. 7-4A**, middle). Most particles represented end or side views of the complex, with few oblique views present (**Fig. 7-4B**). Local resolution varied from 3.6 Å in the GroEL equatorial

domains to worse than 6 Å for bound substrate (Fig. 7-8C). Global resolution was estimated at 3.8 Å (**Fig. 7-4D**).

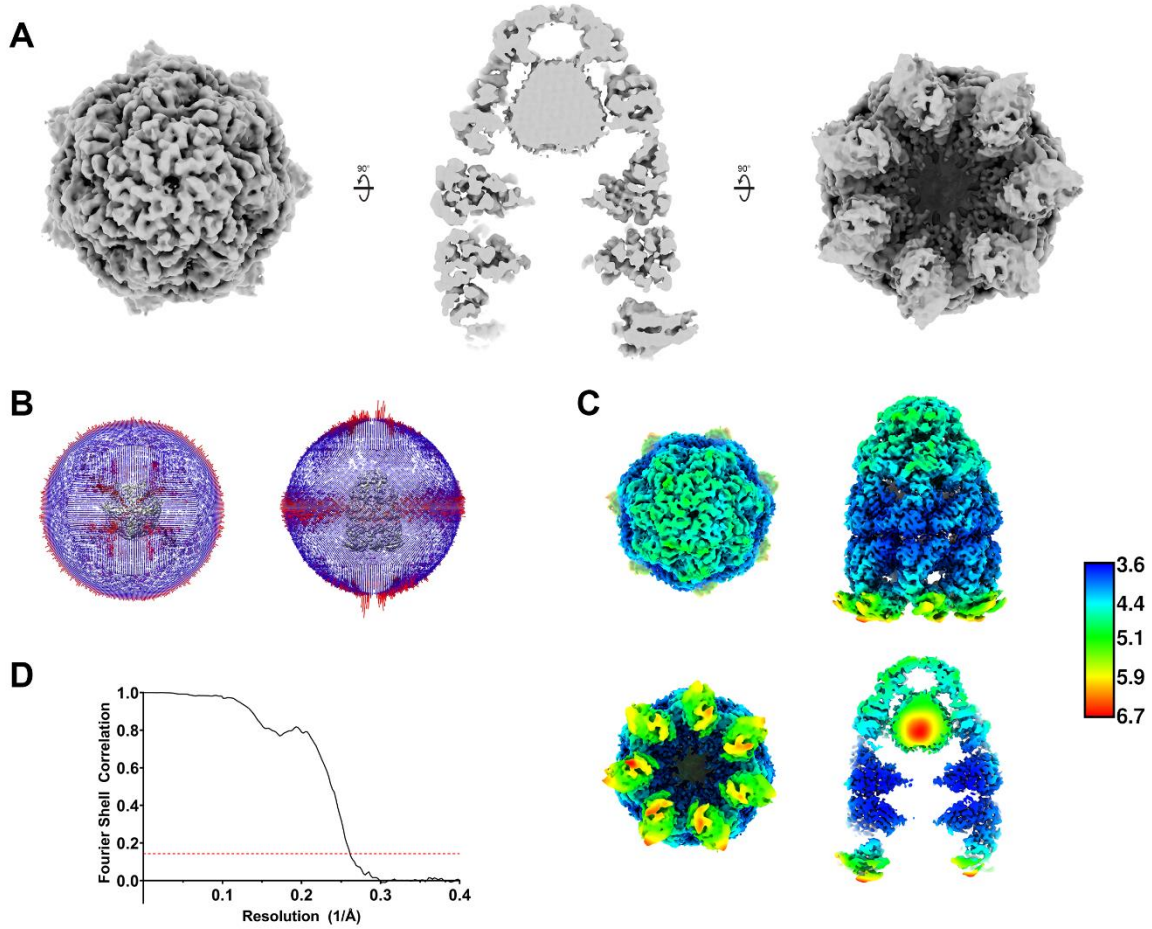

Figure 7-4. Reconstruction of GroEL-GroES-Rubisco. (**A**) Reconstruction of substrate-bound GroEL-GroES viewed from above the *cis* ring (left), a central slice of a side view (middle), and from above the trans ring (right). (**B**) Orientation distribution plot. (**C**) Local resolution of cryoEM map. Colour scale shown by the right bar, with units in Å. (**D**) Fourier shell correlation (FSC) plot for the calculated half-maps; global resolution was 3.8 Å.

7.1.5. Molecular model of GroEL-GroES-Rubisco-ADP·AlF$_x$

To derive an atomic model of GroEL-GroES with encapsulated Rubisco, we first performed rigid body fitting of the published crystal structure of GroEL-GroES-ADP·AlF_3 (PDB: 1SVT) (**Fig. 7-5A**). The *cis* ring of the crystal structure was a good fit for our cryoEM map, with a CC value of

0.80 (**Fig. 7-5B**). However, the *trans* ring apical domains in our reconstruction were in a different, more open conformation than the published crystal structure and the fit was poor (CC = 0.59) (**Fig. 7-5C** and **Fig. 7-6A**). We replaced the model of the *trans* ring with that from our cryoEM structure of GroEL-Rubisco-ADP·BeF_x (**Fig. 7-6B**). The fit improved (CC = 0.82) and required only minor optimisation in ISOLDE to maximise the fit (CC = 0.87) (**Fig. 7-6C**). After combining the *cis* and *trans* ring models, we further optimised the model in ISOLDE to produce a final structure (**Fig. 7-7**). Selected areas of the map and model are shown in **Fig. 7-8**.

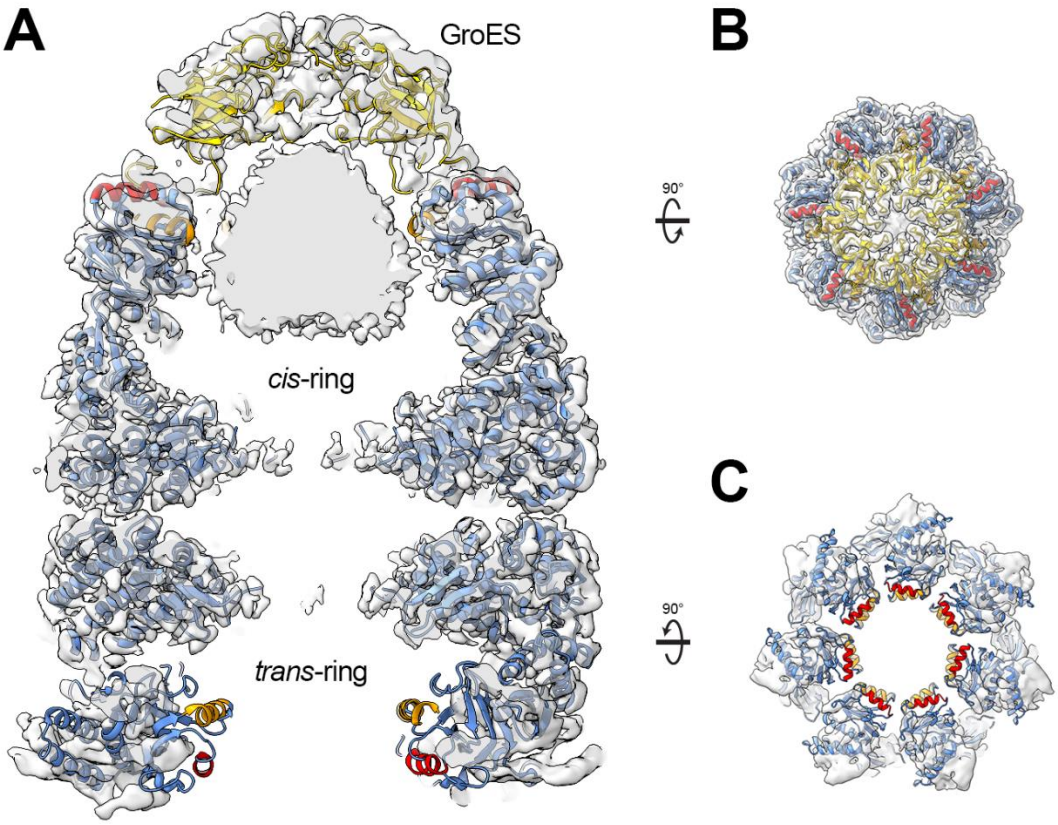

Figure 7-5. Comparison of the GroEL-GroES-ADP·AlF_3 crystal structure and the cryoEM map of GroEL-GroES-Rubisco-ADP·AlF_x. (**A**) Central-slice side view, (**B**) *cis* ring view, and (**C**) *trans* ring view of the result of a rigid body fit of PDB model 1SVT into the cryoEM reconstruction. Density for non-native Rubisco is left unmodelled.

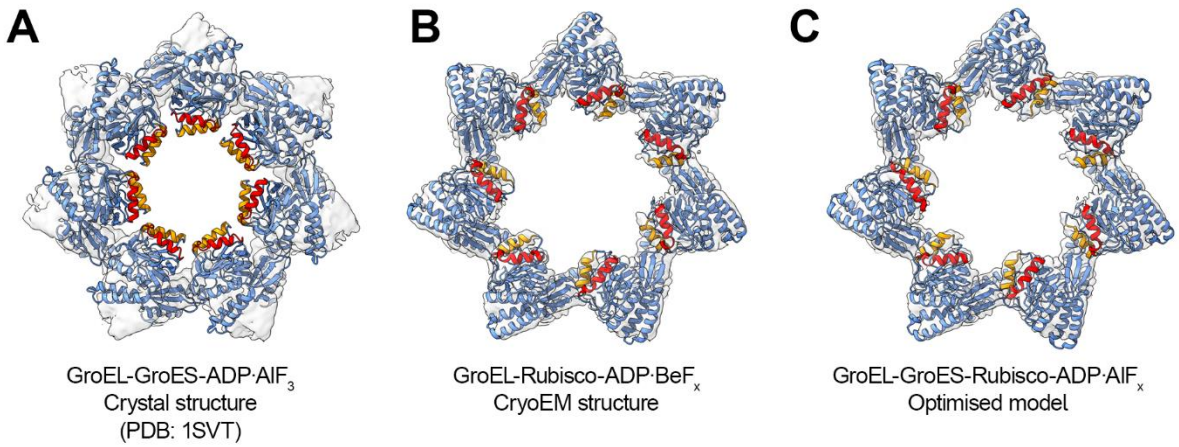

Figure 7-6. Building the *trans* ring model. Comparison of the *trans* ring models of the (**A**) GroEL-GroES crystal structure, (**B**) our cryoEM structure of GroEL-Rubisco-ADP·BeF_x and (**C**) the optimised model of GroEL-GroES-Rubisco-ADP·AlF_x.

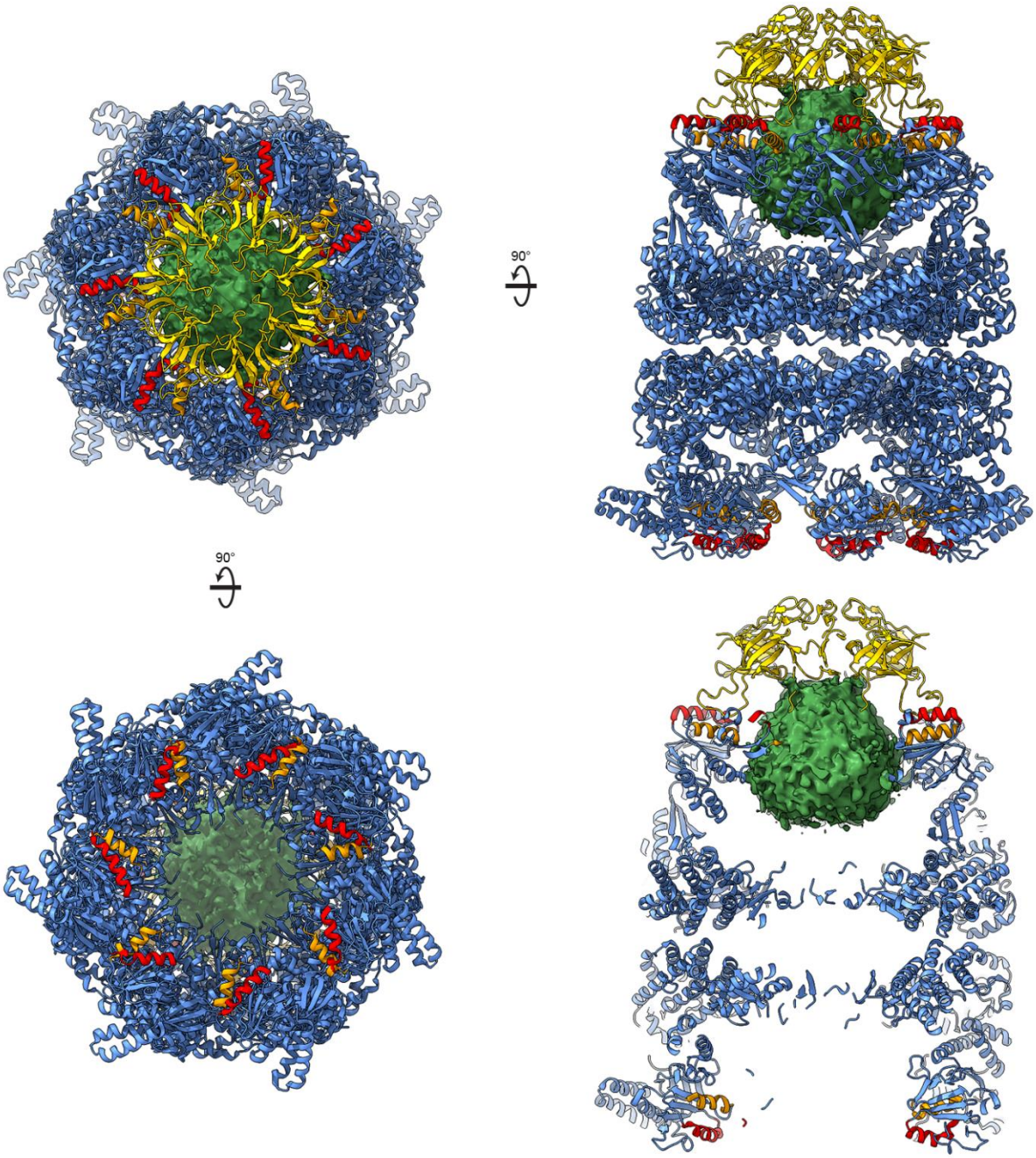

Figure 7-7. CryoEM structure of GroEL-GroES-Rubisco-ADP·AlF$_3$. Structure of the *cis* ring was optimised using the crystal structure of GroEL-GroES (PDB: 1SVT). The structure of the *trans* ring was optimised starting from our structure of GroEL-Rubisco-ADP·BeF$_x$. Density for non-native Rubisco is left unmodelled (green). Atomic model is shown as a ribbon diagram and is coloured as follows; GroEL (blue), helix H (red), helix I (orange), GroES (yellow).

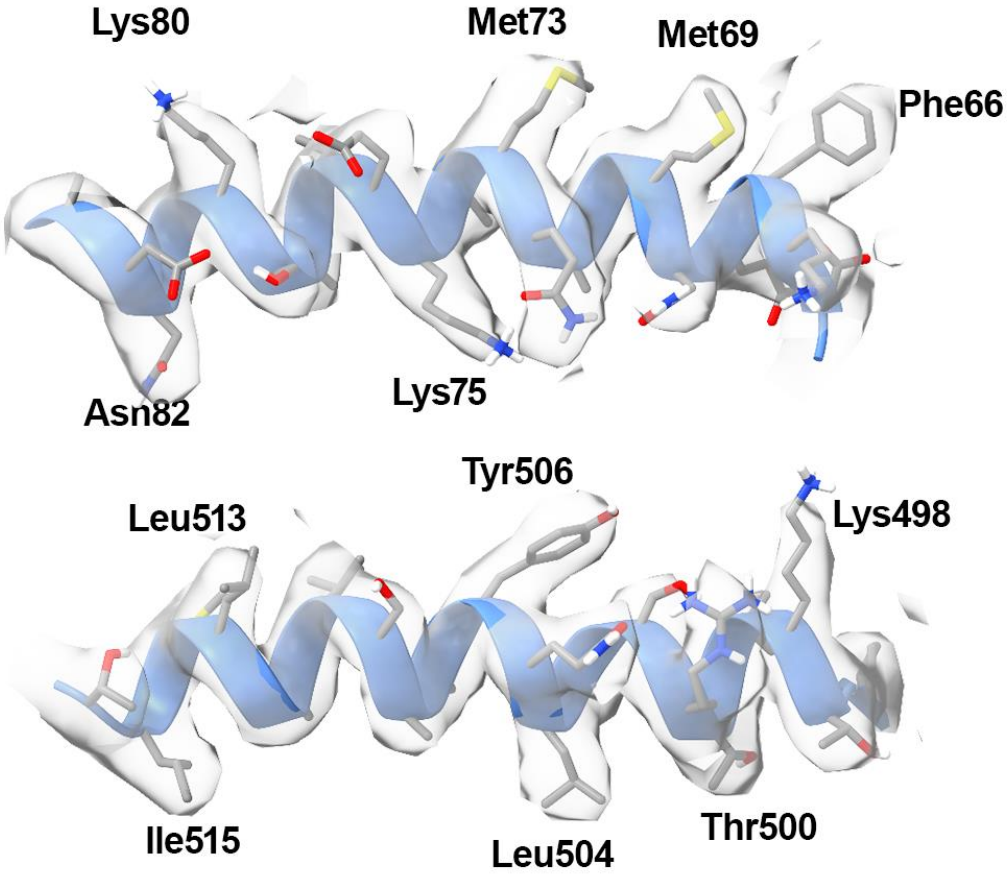

Figure 7-8. CryoEM map quality. Isolated density of two GroEL helices illustrating the map-model fit.

7.1.6. Nucleotide state of *cis* and *trans* ring subunits

The previosuly published crystal structure of GroEL-GroES-ADP·AlF_3 (PDB: 1SVT) has seven ADP·AlF_3 bound in the GroEL *cis* ring, and a nucleotide-free *trans* ring. We examined the nucleotide binding site of each GroEL subunit in our cryoEM reconstruction. Strong density representing ADP was seen in all fourteen GroEL subunits. We observed continuous density between the D87 sidechain and the ADP β-phosphate. At a local resolution of 3.6 Å, we were able to confidently model ADP and the phosphate oxygen-coordinating metal, Mg^{2+}. ADP bound in the *cis* ring showed additional density that we attributed to the ATP γ-phosphate analogue AlF_3 (**Fig. 7-9A**). Nucleotide bound in the *trans* ring lacked this additional density and was modelled as ADP without AlF_3 (**Fig. 7-9B**). *Trans* ring subunits showed the presence of the second

coordinating metal, K^+ (**Fig. 7-9B**). GroEL requires K^+ to hydrolyse ATP and a published crystal structure has confirmed this position as the K^+ binding site (Kiser et al., 2009). AlF_3 is a γ-phosphate analogue of the transition-state of ATP hydrolysis. At this point along the ATP hydrolysis reaction coordinate, K^+ has presumably fulfilled its catalytic role and is not required. In contrast, our high-resolution structure of GroEL-Rubisco-ADP·BeF_x showed K^+ bound in the *cis* ring, but not in the *trans* ring (see **Section 6.3.8**). This is consistent with a pre-hydrolysis ATP analogue. We saw no appreciable variability in the nucleotide density among subunits within the same ring.

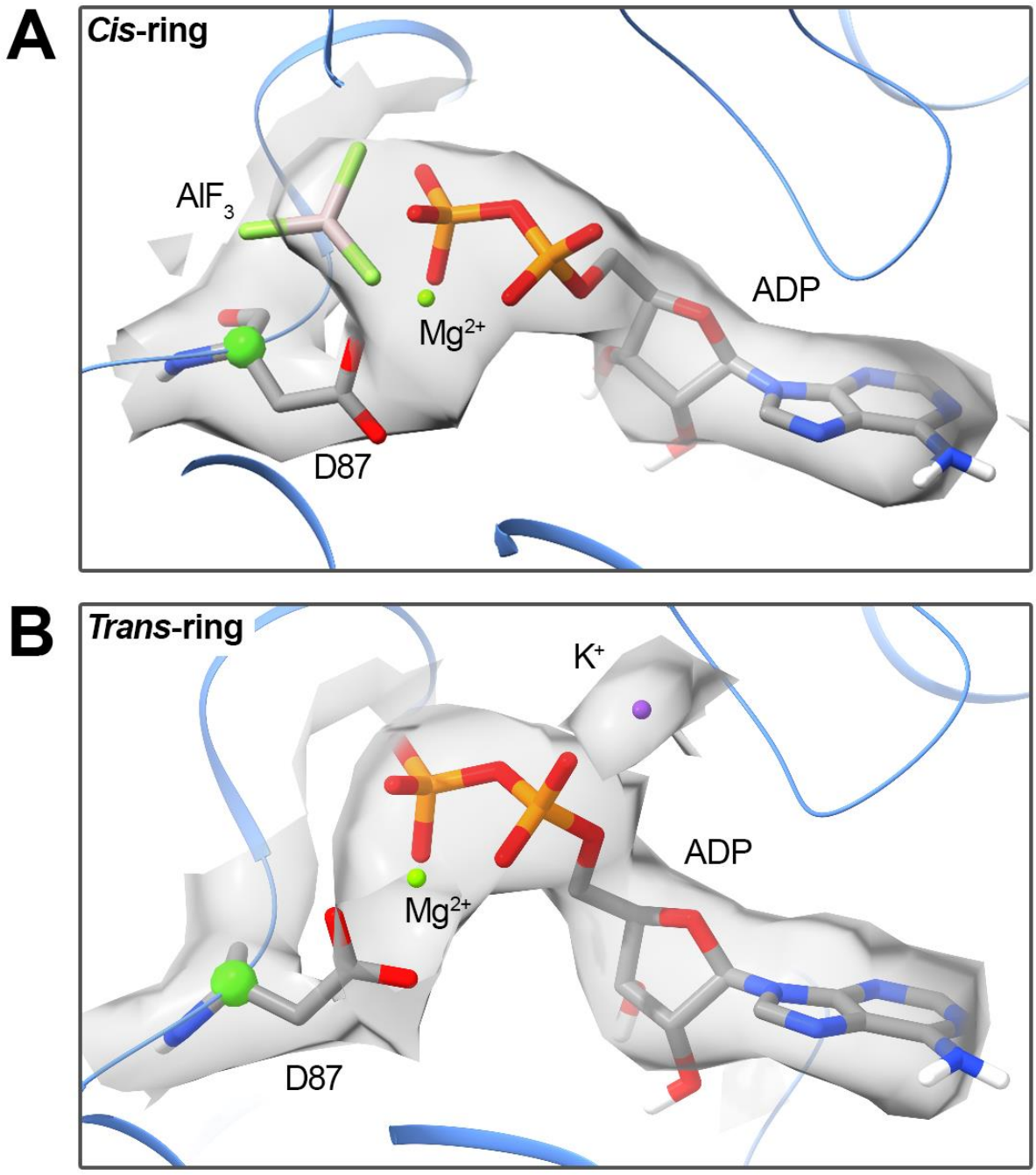

Figure 7-9. Conformation of bound nucleotide in GroEL *cis* and *trans* rings. ADP, Mg^{2+} and K^{+} were placed using ISOLDE and optimised. AlF3 was placed and optimised using Coot. The sidechain of residue D87 is shown. Remaining GroEL structure is represented by a blue ribbon diagram.

7.1.7. Substrate interactions with GroEL-GroES

We next examined the interactions between encapsulated Rubisco and the GroEL-GroES cavity walls (**Fig. 7-10**). Substrate filled the upper two-thirds of the GroEL *cis* cavity. Interactions were observed with cavity-facing residues of GroEL subunits. Surprisingly, we also observed interactions between non-native Rubisco and GroES. The strongest GroEL-substrate interactions involved residues F281 and K226 (**Fig. 7-10A**). Weaker GroEL-substrate interactions were resolved at lower density thresholds and appeared to involve N229 and E255 (not shown). Additionally, we observed strong interactions between substrate and the GroES reside Y71 (**Fig.7-10B**). The Y71 residues of GroES form a hydrophobic collar on the ceiling of the *cis* cavity and may be important for the folding of some GroEL substrates (Wang et al., 2002).

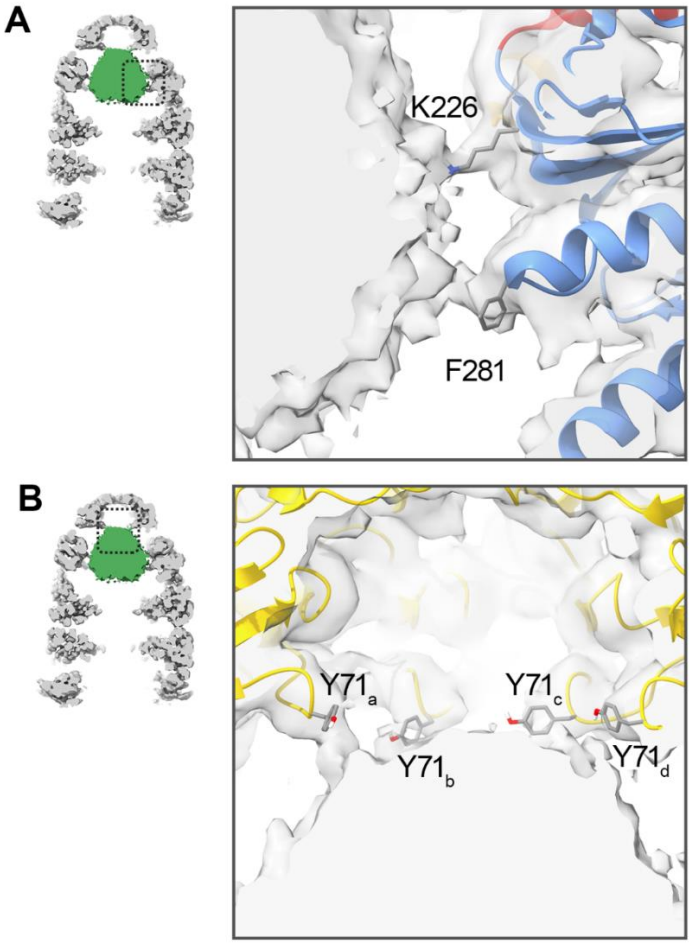

Figure 7-10. Interactions between GroEL and encapsulated Rubisco. (**A**) Interactions between GroEL apical domains and Rubisco. The sidechains of F281 and K226 are labelled. (**B**) Interaction between GroES and Rubisco, with Y71 sidechains shown.

7.1.8. Further image processing revealed substrate density similar to a natively folded Rubisco monomer

The interactions of encapsulated Rubisco appeared to be nearly identical for all GroEL-GroES subunits. The volume of the Rubisco density (~73,000 A^3) accounted for >60 kDa of protein mass, too large for a Rubisco monomer. During image processing, alignment and classification steps were likely dominated by the GroEL-GroES signal. This plausibly could have led to an incorrect averaging of the encapsulated Rubisco density. The substrate density likely represented an ensemble of orientations and conformations of Rubisco inside the GroEL-GroES cavity. We investigated this through further image processing. We first attempted to used particle subtraction in Relion to subtract the GroEL-GroES signal from the 31k GroEL-GroES-Rubisco particles. In principle this would leave only the Rubisco density, allowing us to refine a structure of a native Rubisco monomer. We experimented with various 3D refinement and classification approaches using the new set of subtracted particles. However, we were not able to resolve any recognisable features of Rubisco. We devised an alternative approach again using masked 3D classification of the encapsulated substrate (**Fig. 7-11**). Several of the resulting classes showed asymmetric density for the Rubisco monomer. The most populated class contained 8k particles (**Fig. 7-11**, class 8). The size and shape of class 8 was somewhat reminiscent of a folded Rubisco monomer (**Fig. 7-12**). We refined particles from class 8 and calculated a reconstruction at a global resolution of 4.6 Å (**Fig. 7-13A, D**). The local resolution varied throughout the map, but was lowest for encapsulated Rubisco (**Fig. 7-13B**). Due to the low number of particles, some views were missing, and the map displayed mild anisotropy (**Fig. 7-13C**). Asymmetric conformations were also observed in other 3D classes (**Fig. 7-11**, classes 5 and 7). However, 3D refinement of these classes

resulted in lower resolution reconstructions with significant anisotropy (not shown). Combining visually similar classes (for example class 5 and class 8) resulted in averaging of the Rubisco density and yielded a similar reconstruction to that shown in **Fig. 7-4**. It is likely that these other 3D classes represented native or near-native Rubisco in different orientations and/or conformations.

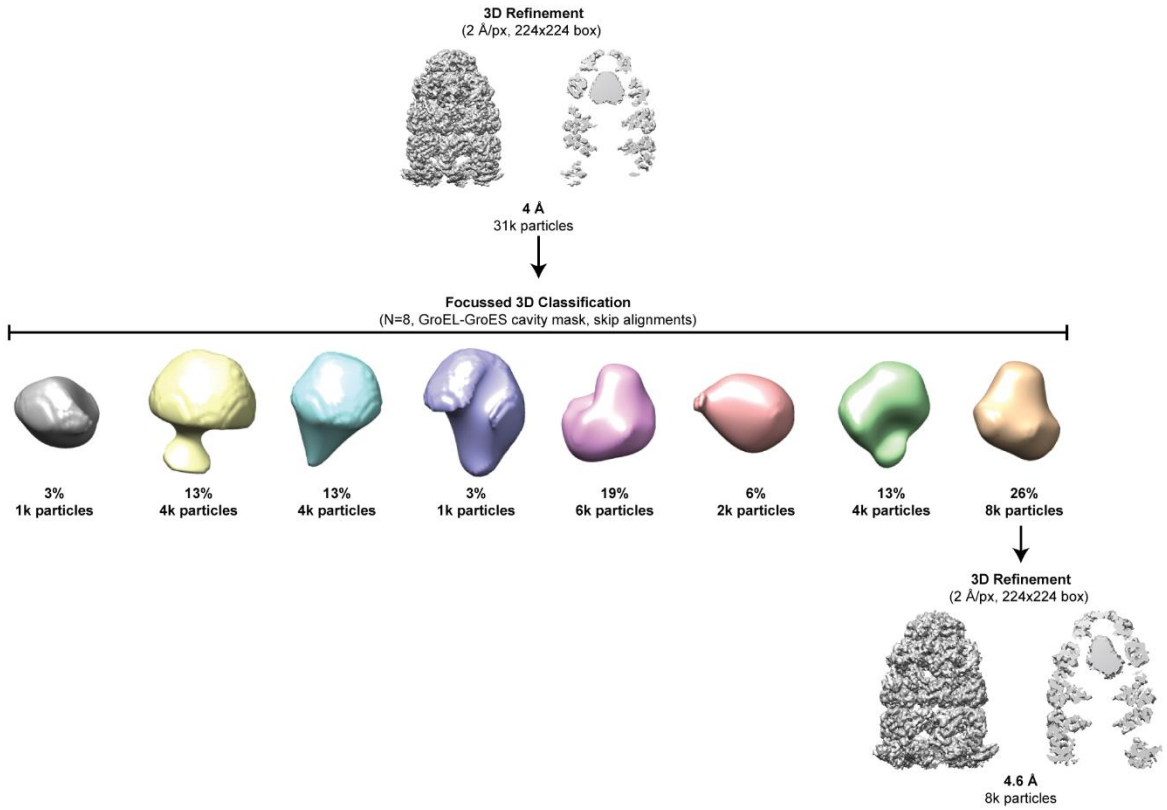

Figure 7-11. Focussed 3D classification of encapsulated Rubisco. Particles showing strong substrate density were subjected to 3D classification without alignments in Relion v3.1. Amplitude contrast was set to 0.3. Classification was performed on a masked region of the map. The ellipsoid mask encompassed the *cis* cavity of GroEL.

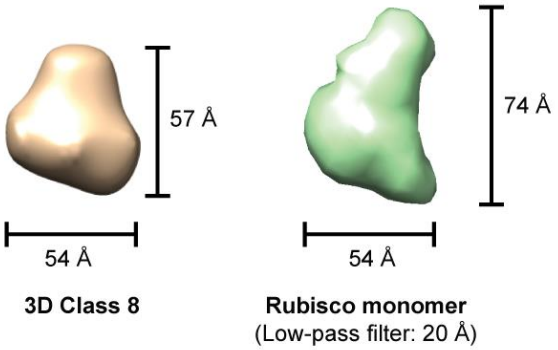

Figure 7-12. Comparison of 3D class 8 and a native Rubisco subunit. Rubisco monomer taken from the crystal structure of Rubisco (PDB: 5RUB, chain A), converted to a volume and low-pass filtered to 20 Å. Distances were measured in UCSF Chimera (Pettersen et al., 2004).

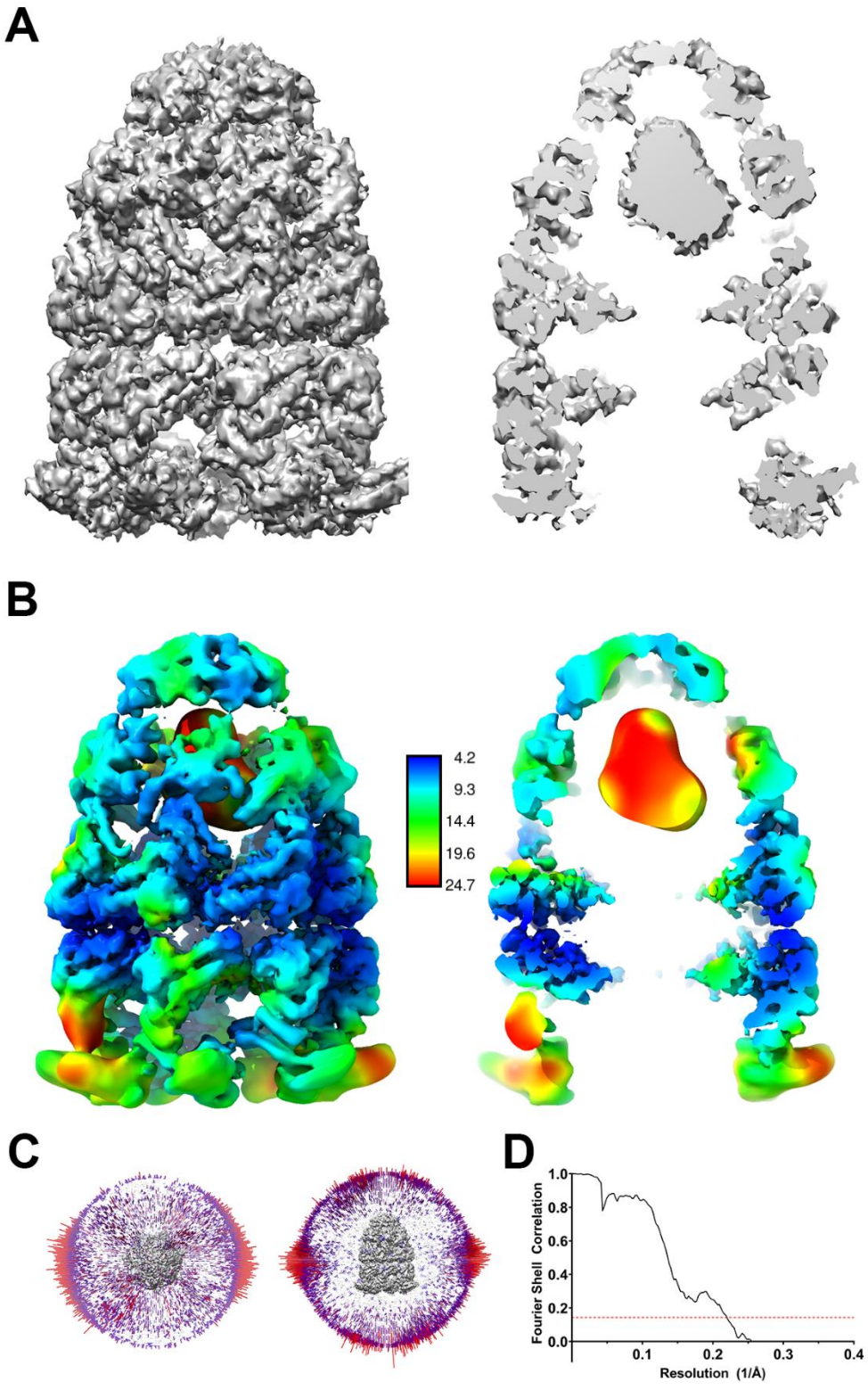

Figure 7-13. CryoEM reconstruction of GroEL-GroES-Rubisco. (**A**) Unsharpened and (**B**) locally filtered maps. Local resolution coloured according to the central bar, with units of Å (**C**) Euler angle plot showing the angular distribution of particles. (**D**) FSC plot of the two half maps. Global resolution was 4.6 Å.

7.1.9. Encapsulated Rubisco folded to a native or near-native state

The volume of the substrate density accounted for 50 – 60 kDa, suggesting it may represent an average of several positions or orientations of Rubisco. We performed manual fitting of a single Rubisco subunit (PDB: 5RUB, chain A) to the density inside the GroEL-GroES cavity (**Fig. 7-14**). We placed the dimerisation interface of the Rubisco monomer against the GroEL-GroES cavity wall. We then used rigid body fitting to optimise the fit (CC = 0.66). Due to the low resolution, the orientation of the Rubisco monomer was ambiguous. Several possible orientations could be modelled into the density. For example, the dimerisation interface (coloured magenta in **Fig. 7-14**) may instead face toward the floor of the *cis* cavity instead of the cavity wall.

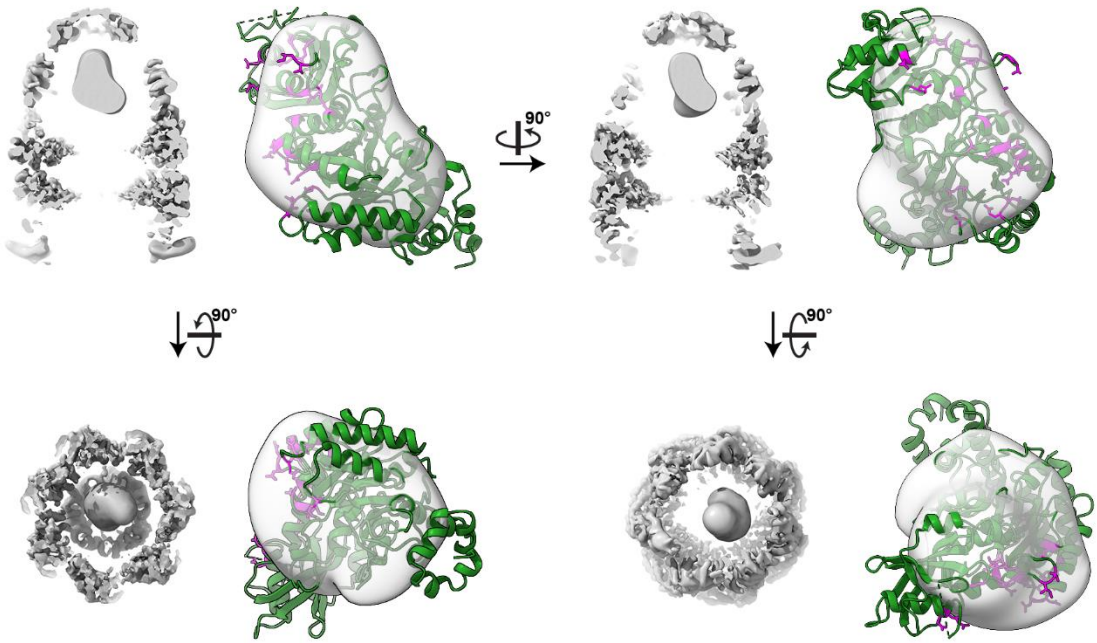

Figure 7-14. Encapsulated near-native Rubisco. Four views of the density representing encapsulated Rubisco. Each quadrant shows a different view. In each quadrant, the image on the

left shows the unmodelled reconstruction, and the image on the right shows an enlarged view of the isolated substrate density with a native Rubisco monomer fitted. Residues involved in inter-subunit contacts in native Rubisco are coloured magenta and shown with visible sidechains.

7.1.10. Non-native Rubisco may have been sequestered by the GroEL C-terminal tails

After identifying GroEL-GroES particles with native (or near-native) Rubisco in the folding chamber, we examined the remaining set of particles for substrate density. We experimented with several different masks targeting different areas of the GroEL-GroES *cis* cavity. By using a mask that encompassed the lower half of the *cis* cavity, close to the C-termini, we were able to resolve additional weak density (**Fig. 7-15**). The C-terminal tails are not typically resolved in crystal structures or cryoEM reconstructions of GroEL. This density might therefore represent non-native Rubisco bound to the C-terminal tails of *cis* ring GroEL subunits.

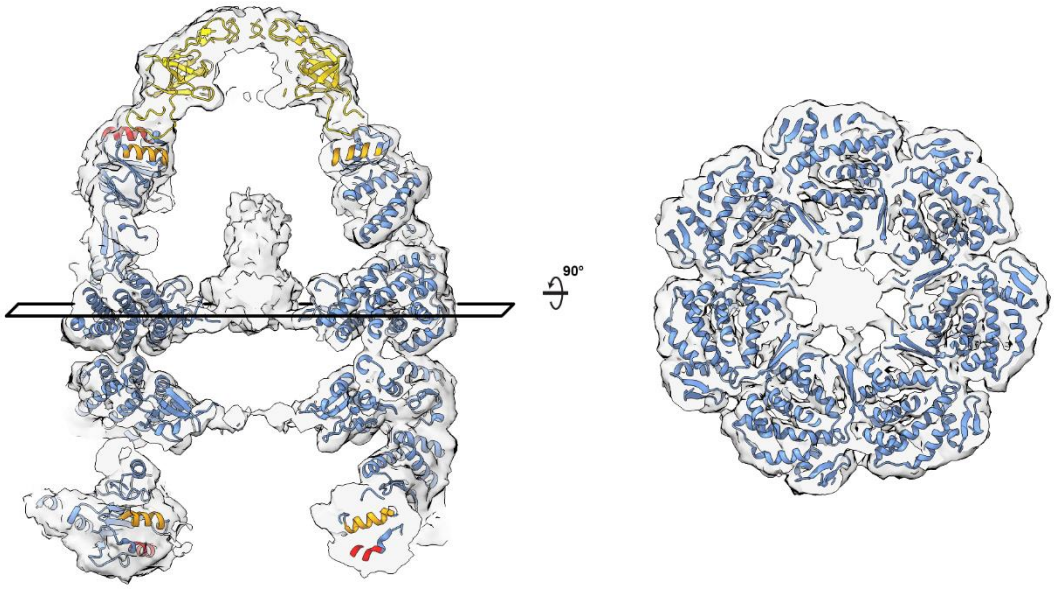

Figure 7-15. Encapsulated non-native Rubisco.

Chapter VIII

8. Discussion

The aim of this project was to investigate molecular mechanisms of assisted protein folding by the GroEL-GroES system. We approached this question by determining cryoEM reconstructions of substrate-bound complexes at different stages of the GroEL-GroES reaction cycle. We first studied substrate binding by forming binary complexes of GroEL bound to non-native Rubisco. Our native mass spectrometry data suggested 1:1 binding between Rubisco monomers and GroEL tetradecamers, in agreement with previous work (van Duijn et al., 2007). Our native PAGE results were ambiguous. This was due to the inexact mass measurements in native PAGE and the presence of two substrate-bound species of GroEL. The additional bands might have represented different occupancies of non-native Rubisco, or different binding modes. We could not rule out the possibility of double binding of non-native Rubisco.

Our initial cryoEM reconstruction of GroEL-Rubisco (from untilted data) showed weak density for bound substrate. We were unable to convincingly resolve specific contacts between GroEL and Rubisco. However, the overall conformation of the non-native substrate was somewhat similar to that seen in a previously published cryoEM reconstruction of GroEL-Rubisco (Natesh et al., 2018). In the reconstruction by Natesh et al. (2018), substrate density accounted for 30 – 35% of the volume of a folded Rubisco monomer. Our initial efforts gave similar results.

Weak and incomplete density for bound non-native substrate was caused by preferred orientation, a technical problem common to cryoEM studies of chaperonins (Elad et al., 2007, Jin et al., 2019 and Klebl et al., 2021). The preferred orientation of GroEL-Rubisco presented a significant obstacle to our high-resolution structure efforts. We attempted to use conventional biochemical and grid-targeted methods to address the preferred orientation but were unsuccessful. Ultimately,

two different approaches led to improved reconstructions of GroEL-Rubisco. The first approach was to collect cryoEM data at a high tilt angle. Tilting provided a broader distribution of views of the complex, but did nothing to alter the interaction of GroEL-Rubisco with the air-water interface. The tilt data set produced two moderate-resolution reconstructions with strong substrate density. The second approach we used was to prepare cryoEM grids using a Chameleon instrument. By lowering the time between sample application and plunge-freezing, we reduced the interaction of GroEL-Rubisco with the air-water interface. The preferred orientation was still present, but to a lesser extent. Enough side and oblique views of substrate-bound complexes were collected to allow for a more isotropic reconstruction. A combined strategy employing both Chameleon and stage tilt could be an interesting avenue to explore for especially challenging samples.

We were unable to answer conclusively how many non-native Rubisco monomers can bind to GroEL. Our cryoEM reconstruction from Chameleon-prepared grids showed density for non-native Rubisco in both GroEL cavities. However, density for the second bound Rubisco was weak, and only visible at a low threshold. Conversely, we did not observe a second bound Rubisco in reconstructions from our tilt data set (even at low threshold values), or in our native mass spectrometry experiments. If a second non-native Rubisco monomer does bind to GroEL, it appears to do so weakly. If non-native Rubisco bound identically to both rings, it would be visible in our cryoEM reconstructions. It is possible that some GroEL particles were misaligned during 3D refinement. If the wrong "up-down" orientation of some GroEL particles had been assigned, it could lead to the presence of Rubisco density in both GroEL cavities. In native mass spectrometry, ionisation of GroEL-Rubisco complexes may have caused weakly bound Rubisco to dissociate, leaving only singly-bound complexes.

The reduced interaction of GroEL-Rubisco with the air-water interface in the Chameleon data sets yielded additional benefits. We resolved more Rubisco density from the Chameleon data set

(99% of a full Rubisco monomer) than from the tilt data set (60%). Substrate density in the Chameleon data set also resembled the size and shape of a folded Rubisco monomer and likely represented a Rubisco folding intermediate or molten-globule state. In contrast, substrate density in the best tilt data set reconstruction was incomplete and not recognisable as a Rubisco monomer. CryoEM grids used in the tilted data set had been prepared with conventional blotting and plunge-freezing methods using a Vitrobot. Non-native Rubisco likely partially denatured during the much longer exposure (at least 5 – 10 s) to the air-water interface. This could explain why we did not observe substrate density protruding from the GroEL cavity in our reconstructions from Vitrobot-prepared grids. Additionally, the grid used in the tilted data set had been prepared with an additional monolayer of graphene oxide. GroEL-Rubisco particles likely adsorbed to both graphene oxide and to the air-water interface. Proteins are thought to be protected from denaturation at the graphene-water interface (compared to the air-water interface). However, this is less likely to be true for a non-native protein. Because we did not observe density for a full Rubisco monomer, it is likely that non-native Rubisco had partially unfolded or denatured in the tilted data set.

Previously published cryoEM reconstructions of GroEL-substrate complexes (including GroEL-Rubisco) did not display density protruding from the GroEL cavity. These reconstructions also did not account for the entire mass of the substrate. Previous works report the following values for the percent mass their cryoEM reconstructions account for; GroEL-MDH (25-40%) (Elad et al., 2007), GroEL-gp23 (54% in binary, 76% in cis ternary) (Clare et al., 2009), GroEL-actin (28%; however, actin is not a true GroEL substrate) (Balchin et al., 2018), and GroEL-Rubisco (30-35%) (Natesh et al., 2018). Most authors presumed the substrate to be partially disordered (Elad et al., 2007, Clare et al., 2009, Balchin et al., 2018 and Natesh et al., 2018). It is likely that the substrate in these studies had instead been at least partially denatured by contact with the air-water interface. All previously published GroEL-substrate structures came from grids prepared by conventional plunge-freezing methods, introducing a long delay between sample application on the grid and

vitrification of the sample. We therefore hypothesize that the use of the Chameleon in our study was key in observing the full mass of a bound non-native substrate.

8.1. Non-native Rubisco bound to GroEL in two distinct modes

Our structural study of GroEL-Rubisco builds on previous work demonstrating that non-native substrates bind multivalently to different arrangements of GroEL subunits (Farr et al., 2000 and Elad et al., 2007). Stringent substrates such as MDH require 3 – 4 consecutive GroEL subunits to mediate full binding (Farr et al., 2000 and Elad et al., 2007). Less stringent substrates such as Rhodanese (33 kDa) require fewer GroEL subunits (Farr et al., 2000). Rubisco, a large stringent substrate, was shown to bind more efficiently to different arrangements of non-consecutive GroEL subunits (Farr et al., 2000 and Natesh et al., 2018).

Our cryoEM reconstructions showed non-native Rubisco bound to GroEL in two different modes. The tilt data set showed Rubisco bound to opposite pairs of GroEL subunits, suspended across the centre of the cavity. This mode of substrate binding was among several studied by Farr et al. (2000), it supported more efficient binding of Rubisco (50.5 kDa) compared to a smaller substrate, MDH (33 kDa). The Chameleon data set showed Rubisco bound strongly to three consecutive subunits on one side of the GroEL cavity. At lower density thresholds, reconstructions from both data sets showed all seven GroEL subunits involved in contacting non-native Rubisco. However only the chameleon data set suggested an interaction with the disordered GroEL C-termini.

8.2. Specific interactions between GroEL and Rubisco

Previous cryoEM studies of GroEL-substrate complexes have been limited to resolutions of 8 – 12 Å. Interactions between non-native substrate and GroEL were typically not well resolved. Taking advantage of recent advances in cryoEM, our reconstructions of GroEL-Rubisco allowed us to identify specific contact points between non-native Rubisco and residues of GroEL subunits.

The network of interacting residues differed slightly between each binding mode of non-native Rubisco. However, both data sets showed that the strongest contacts were formed with helix I of GroEL, consistent with previous cryoEM studies of different GroEL-substrate complexes (Elad et al., 2007, Clare et al., 2009, Balchin et al., 2018 and Natesh et al., 2018). Many of the residues involved in contacting non-native Rubisco were those identified in the original mutational studies of GroEL (Fenton et al., 1994). Additionally, Rubisco interacted strongly with Met-267, Arg-268 and Ile-270, three residues located at or near the C-terminus of helix I. The importance of these residues in substrate binding/folding is less well characterised. Structural evidence for the role of Arg-268 and Ile-270 in mediating substrate interactions comes from the crystal structure of GroEL bound to a short (12-residue) polypeptide, called "strongly-binding peptide" (SBP). SBP made van der Waals contacts to Arg-268 and Ile-270, as well as hydrogen bonding to Arg-268 (Wang and Chen, 2003). In all of our reconstructions, the single strongest GroEL-substrate interaction involved Arg-268. Substrate binding to GroEL is typically characterised by hydrophobic interaction. Strong binding to Arg-268 might be a feature of stringent substrates, such as Rubisco. Additional studies are needed to explore the relative importance of Arg-268 over other residues. We were surprised to see a strong interaction between non-native Rubisco and Met-267. This residue has not previously been implicated in substrate binding, but in allosteric communication within and between GroEL subunits (Kass and Horovitz, 2002). However, an exact allosteric role for Met-267 is not known. Substrate binding to Met-267 might be involved in the transmission of negative cooperativity to the opposite GroEL ring, perhaps in conjunction with binding to Arg-268. Substitution of these residues, followed by additional binding studies could explore this further.

Residues close to the C-terminus of helix I (including Val-264, Met-267, Arg-268, Ile-270) formed the strongest interactions with non-native Rubisco. This region lies deeper in the GroEL cavity than the N-terminus of helix I, or any part of helix H. Following the initial capture by helices H and I, non-native substrate might be directed deeper inside the cavity, strengthening the interaction

with GroEL, and preventing escape. Progressively deeper binding of non-native substrate might promote an initial interaction with the disordered C-terminal tails of GroEL. The interaction with the C-termini strengthens following the ATP-induced conformational changes of GroEL (discussed later).

In our experiments, the time between formation of GroEL-Rubisco binary complexes and cryoEM grid preparation was on the scale of minutes, far longer than the physiologic time the complex would exist *in vivo*. Therefore, our reconstruction from the Chameleon data set might represent the maximum extent to which a folding intermediate is (partially) encapsulated by GroEL prior to ATP and GroES binding.

8.3. On the mechanism of negative cooperativity for substrate binding

Non-native Rubisco has been shown to bind to GroEL with 1:1 stoichiometry, exerting negative cooperativity on the opposite ring and inhibiting further binding of Rubisco (van Duijn et al., 2007). Other substrates, such as gp5 (47 kDa), gp23 (56 kDa) and MDH (33 kDa) can bind to both rings simultaneously (van Duijn et al., 2007). Our native mass spectrometry results suggested the same 1:1 binding for Rubisco. However, our results from cryoEM were less conclusive. Reconstructions from our tilt data showed Rubisco bound in only one GroEL ring. Any potential substrate density in the opposite ring was of a similar strength to noise. We therefore concluded that our tilt data reconstructions exhibited true 1:1 binding of Rubisco to GroEL. Our Chameleon data suggested that Rubisco bound to both rings simultaneously, but with different affinities. Why did we observe double binding of Rubisco for the Chameleon data set, but not the tilt data set? Prior to freezing, GroEL-Rubisco particles in the tilt dataset had been confined to a thin layer of buffer, and subjected to the air-water interface (or the graphene-water interface) for a longer period of time (5 – 10 s) than in the Chameleon data sets (100 ms or 1 s). Prolonged contact with the air-

water interface could have stripped GroEL-Rubisco particles of the second, weaker bound Rubisco, leaving only particles with strongly bound Rubisco. This could explain the observation of singly bound Rubisco in the tilt data set, and doubly bound Rubisco in the Chameleon data set.

We only observed extra density for the R430-E434 inter-ring salt bridge in our tilt data set. If this salt bridge is involved in maintaining the negative cooperativity of Rubisco binding, its absence could imply that the negatively cooperative effect for Rubisco binding had somehow been overcome in our Chameleon data set reconstruction. The salt bridge may only be present when one Rubisco is bound. R430-E434 has previously been implicated in the allosteric mechanism of inter-ring negative cooperativity for ATP binding (Skjaerven et al., 2012). But a similar role in the negative cooperativity for substrate binding has not been reported. The residue E434 is known to be important in general inter-ring communication and is thought to form a salt bridge to K105 of opposite GroEL subunits (Sot et al., 2009). Rubisco binding might switch the salt bridge from K105-E434 to R430-E434, perhaps transmitting negative cooperativity and preventing substrate binding in the opposite ring. However, it is not clear why this salt bridge was not resolved in our Chameleon data set. Previous work has suggested that the structural basis for the negative cooperativity of Rubisco binding lies in a narrowing of the opposite GroEL ring (van Duijn et al., 2007, Chen et al., 2013 and Natesh et al., 2018). However, we did not see any significant structural changes in the opposite ring to support this. Both ring openings had a diameter of ~50 Å. Therefore, the mechanism for negative cooperativity of substrate binding remains unclear.

8.4. GroEL-ATP simultaneously sequesters non-native substrate and recruits GroES

We aimed to determine the effects of ATP binding on the conformation of GroEL-substrate complexes. Our BN-PAGE experiments had shown marginally slower migration of ATP-bound GroEL-Rubisco compared to nucleotide-free or ADP-bound complexes. Binding of a maximal

fourteen ATP molecules would impart a mass increase of ~7 kDa, or 0.8%. Such a small increase would not likely be visible by BN-PAGE. Therefore, the slower migration observed for ATP-bound GroEL-Rubisco was likely instead primarily caused by the conformational changes of GroEL. This effect was identical when ATP was replaced by the ATP analogues ADP·BeF_x or ADP·AlF_x. We determined preliminary cryoEM structures of GroEL-Rubisco bound to ATP, ADP·BeF_x or ADP·AlF_x and saw similar structural features in each reconstruction. The two ATP analogues imparted the same structural changes in GroEL as ATP did, but seemed to make the complex less prone to partial denaturation at the air-water interface (presumably because the complex is stalled and does not continue to cycle ATP). We therefore decided to pursue a high-resolution structure of GroEL-Rubisco bound to the ATP ground state analogue, ADP·BeF_x.

Our structure of GroEL-Rubisco-(ADP·BeF_x)$_7$-ADP_7 offers the first view of a distinctly asymmetric wild-type GroEL ring. The degree of asymmetry in the substrate-occupied ring is particularly striking. It was immediately clear that different GroEL subunits could fulfil discrete roles in this intermediate complex. The apical domains of four GroEL subunits sequestered non-native Rubisco in the *cis* cavity. Compared to the nucleotide-free complex, the interaction between GroEL and Rubisco was strengthened by the involvement of the C-terminal tails of all seven GroEL subunits. The conformation of substrate-contacting subunits was also stabilised by the novel equatorial-apical domain salt bridge R58-E209. This salt bridge was absent from binary complexes. Formation of this salt bridge might be promoted by the combined presence of non-native substrate and the ATP γ-phosphate, mimicked in our structure by BeF_x. The R58-E209 salt bridge was not observed in the published cryoEM structures of GroEL-ATP (Clare et al., 2012). However, this was likely due to the lower resolution of the cryoEM maps.

The remaining three GroEL subunits adopted an extended conformation, presumably to promote recruitment of GroES. Our structure therefore represents a transient, intermediate conformation of GroEL, adopted in response to ATP binding, and immediately prior to GroES binding and

substrate encapsulation. In this state, substrate is prevented from escaping and aggregating in solution. Interestingly, the two functions of GroEL subunits (substrate binding and GroES recruitment) are spread evenly around the GroEL ring. Or in other words, there is not a "substrate-binding half" and a "GroES-binding half" to a GroEL ring. This was somewhat surprising as substrates of GroEL have been shown to fold more efficiently when their initial binding occurs to 3 – 4 consecutive GroEL subunits (Farr et al., 2000). Only two of the substrate-contacting GroEL subunits were consecutive in the ADP·BeF_x-bound state. This weakening of the interaction between GroEL apical domains and substrate might be offset by the stronger association with the C-terminal tails.

Previous studies have speculated that GroES is initially recruited by 1 – 2 raised GroEL subunits that do not contact non-native substrate (Elad et al., 2007 and Clare et al., 2012). Significant asymmetry within a single GroEL ring has only been previously observed in the crystal structure of the double mutant GroELΔD83A/ΔR197A bound to ADP (Fei et al., 2013). In this mutant, two inter-subunit salt bridges were removed, effectively detaching adjacent apical domains. The freed apical domains adopted different conformations that were similar to those observed in the symmetric GroEL rings by Clare et al. (2012).

Studies of GroEL bound to alternate ATP analogues, such as ATPγS or AMP-PNP, fail to form folding active complexes and do not show the same large-scale conformational changes in GroEL (Boisvert et al., 1997, Rye et al., 1997 and Chaudhry et al., 2003). This is likely related to the critical role of the ATP γ-phosphate.

We observed only one asymmetric arrangement of *cis* ring subunits in GroEL-Rubisco-ADP·BeF_x. Our cryoEM structure might therefore represent a favourable energetic minimum that provides a balance between its two functions (i.e. substrate binding and GroES binding). This intermediate structure might be substrate independent, or different substrates might yield different arrangements of GroEL subunits. For example, it is conceivable that smaller, less stringent

substrates such as Rhodanese would only require two GroEL subunits to maintain contact while the other five adopt an elevated conformation to recruit GroES.

We saw no direct structural evidence for a "forced unfolding" of bound substrate upon binding of ATP. However, a mechanism can be inferred. The four substrate-contacting subunits are arranged in such a way that non-native Rubisco is suspended across the centre of the cavity, rather than localised to one side, such as in our nucleotide-free GroEL-Rubisco structure. Additionally, non-native Rubisco formed strong contacts to the disordered C-termini of GroEL. When the four substrate-bound GroEL subunits transition to the GroES-bound conformation, the distances between opposite apical domains approximately doubles. Additionally, the distance between GroEL apical domains and C-termini increases. Bound non-native Rubisco might undergo a moment of rapid conformational expansion as it becomes stretched between GroEL apical domains and the C-termini. Our structure might therefore represent a non-native substrate that is primed for forced unfolding. Upon binding of GroES to all seven GroEL subunits, helices H and I are buried, and Rubisco would be released into the now enclosed *cis* cavity, where it would undergo compaction and folding. This model is partly supported by FRET studies of GroEL-Rubisco (Lin et al., 2013). Compared to our reconstructions of nucleotide-free GroEL-Rubisco and the GroES-bound complex, we could only account for ~50% of the Rubisco volume in our reconstruction of GroEL-Rubisco-ADP·BeF_x. It is not entirely clear where the remaining volume would reside. The most likely explanation is that non-native Rubisco protrudes outside the *cis* cavity in a disordered state. This could explain why substrate density was unresolved even at low density thresholds.

8.5. Rubisco folded to a native-like state inside the *cis* cavity of GroEL-GroES

CryoEM and image processing of GroEL-GroES-Rubisco identified a subset of 31k particles that displayed strong density for encapsulated Rubisco and revealed its interactions with GroEL-GroES. Further classification yielded a smaller subset of ~8k particles that displayed a native-like size and shape for the Rubisco monomer, possibly a molten globule with non-native tertiary structure. This conformation of Rubisco interacted less strongly with GroEL-GroES, suggesting that this state represents a native or native-committed Rubisco monomer, ready to exit the GroEL-GroES cavity. The remaining ~23k particles likely contained Rubisco in different orientations and conformations. These could include off-pathway folding intermediates that require additional rounds of binding and encapsulation.

The previously published cryoEM structure of GroEL43Py398A-GroES-Rubisco-ATP showed density for non-native Rubisco in the *cis* cavity (Chen et al., 2013). However, GroEL43Py398A-GroES is not a fully folding-active complex. The interaction between GroEL43Py398A and Rubisco was likely altered by the presence of the pyrene dye molecules, attached covalently to Cys-43 of the GroEL equatorial domains. Our reconstruction does not suggest an interaction between encapsulated Rubisco and the equatorial domains of GroEL. Instead, we observed a native-like Rubisco suspended in the upper region of the GroEL cavity, making only weak interactions with the cavity wall. Encapsulated native-like Rubisco interacted with several residues of the GroEL folding chamber. The strongest interaction was with Phe-281, one of three hydrophobic GroEL side chains that face the *cis* cavity (the other two residues are Leu-309 and Tyr-360). Rubisco interacted with 3 – 4 Phe-281 residues of consecutive GroEL subunits. The GroEL mutant F281D supports binding of non-native substrate but exhibits decreased ATPase activity, reduced folding, and aggregation of substrate protein upon release from the *cis* cavity (Fenton et al., 1994). This suggests an active role for Phe-281 in GroEL-assisted folding. An interaction between

encapsulated non-native Rubisco and Phe-281 was also proposed by Chen et al. (2013), but not confidently resolved in their cryoEM reconstruction. The interaction between encapsulated Rubisco and GroES Tyr-71 was comparatively weaker, although GroES Tyr-71 has been implicated in the folding of endogenous GroEL substrates (Wang et al., 2002).

Previous work has shown that Rubisco folds to its native state while inside the GroEL-GroES *cis* cavity, but that only around 5 – 10% of Rubisco monomers reach their native, assembly-competent state during each round of binding and encapsulation (Brinker et al., 2001). After each round, both native and non-native Rubisco monomers are released from GroEL. Non-native Rubisco rapidly rebinds to GroEL, while native Rubisco goes on to form an active dimer with other Rubisco monomers. If re-binding of non-native Rubisco was blocked, the released monomers rapidly aggregated in solution (Brinker et al., 2001). When rebinding was permitted, Rubisco was recaptured by GroEL. Folding then proceeded as normal through additional rounds of GroES and ATP-dependent encapsulation and release. Under our experimental conditions, GroES could not dissociate once bound to GroEL due to the presence of non-hydrolysable ADP·AlF_3. Our cryoEM study therefore mimicked a single round of binding and encapsulation, but not release. The study by Brinker et al. (2001) suggested that Rubisco monomers should have continued to fold inside the *cis* cavity, independent of ATP and GroES-cycling (one important distinction is that Brinker et al. (2001) studied a single-ring GroEL). Additionally, ADP·AlF_x has been shown to form folding-active single-ring and double-ring GroEL-GroES complexes with the substrate Rhodanese.

In our GroEL-GroES-Rubisco-ADP·AlF_x data set, we only observed native-like Rubisco monomers in ~4% of GroEL-GroES particles (31k particles showed substantial Rubisco density, out of 814k particles). This suggested that the majority of Rubisco monomers had failed to reach a native-like fold, or that substrate binding/encapsulation had been inefficient. We identified a subpopulation of GroEL-GroES particles with additional density at the base of the *cis* cavity.

However, this could also be attributed to the C-terminal tails of GroEL, rather than to non-native Rubisco. The low efficiency of Rubisco encapsulation and folding can also be partly explained by our experimental conditions. Occupancy of Rubisco in binary complexes was not 100%. Additionally, 50% of the Rubisco bound to GroEL would likely have dissociated upon ATP binding to the opposite (substrate-free) ring. Native PAGE showed evidence of substrate dissociation upon the addition of ATP or ATP analogue. This would have reduced the number of substrate-containing GroEL-GroES complexes by half. One way to avoid this is with the use of a single-ring mutant of GroEL. Single-ring GroEL can mediate near-full recovery of substrate activity, however this precludes the study of inter-ring communication.

8.6. The 7-fold symmetric *trans* rings of GroEL-ADP·BeF_x and GroEL-GroES-Rubisco-ADP·AlF_x showed an open conformation of GroEL

The open conformation of *trans* ring GroEL subunits in our GroEL-ADP·BeF_x and GroEL-GroES-ADP·AlF_x structures appeared to disfavour substrate binding. The conformation we observed was similar to the Rs-open and Rd-open states described in Clare et al. (2012). However, our structures differed slightly in the elevation and orientation of the apical domains. In our structure, *trans* ring apical domains had undergone additional elevation and counterclockwise rotation so that helices H and I faced toward the sides of adjacent GroEL apical domains. Non-native substrates bind most efficiently to multiple consecutive GroEL subunits. However, in our structure of the *trans* ring, the continuous hydrophobic collar formed by helices H and I is disrupted. We saw no hint of substrate density in the *trans* ring of either complex, suggesting that this conformation seems to preclude substrate binding. We also did not observe alternate *trans* ring conformations. This conformation might therefore serve to prevent binding of additional substrate until complete hydrolysis of all seven ATP molecules in the *cis* ring.

The open conformation of the *trans* ring has not been observed in any of the previously published structures of GroEL-GroES, even when the same ATP analogues were used. These include the cryoEM structure of GroEL-gp31-gp23-ADP·AlF_x (Clare et al., 2009), the cryoEM structure of GroEL43Py398A-GroES-Rubisco-ATP (Chen et al., 2013), and the crystal structure of GroEL-GroES-ADP·AlF_x (Chaudhry et al., 2003). In each of these published structures, the *trans* ring of GroEL was devoid of bound ADP and observed to be in the T-state, the same state seen in nucleotide-free GroEL. The open conformation we observed was therefore likely due to the presence of ADP in the *trans* ring. However, it is not clear what caused ADP to remain bound in the *trans* ring of our structures and not of those in previously published structures. Concentrations of nucleotide and BeF_x/AlF_x were similar or identical to previous studies (in the range of 3 – 5 mM). The ratio of ADP·BeF_x to ADP may be an important determinant. The main difference in our study was the combined presence of non-native Rubisco, ADP·AlF_x and wild-type GroEL and GroES (as opposed to mutants or homologues). The GroEL-D398A mutant used by Chen et al. (2013) might have perturbed GroEL interactions with the ATP γ-phosphate, rendering the complex unable to transmit an allosteric signal to the *trans* ring. The use of gp31 as a GroES homologue by Clare et al. (2009) might also have important implications, as GroES can modulate GroEL ATPase activity (Todd et al., 1994). The use of gp23 as a substrate may also represent an important difference, gp23 can bind both GroEL rings simultaneously (Clare et al., 2009), while Rubisco binds one ring at a time (van Duijn et al., 2007). Could our *trans* ring open conformation be related to the presence of non-native Rubisco in the *cis* cavity? We cannot answer this definitively. If the *trans* ring conformation was related to the presence of non-native Rubisco in the *cis* ring, we would expect to see multiple *trans* ring conformations (including the apo-like T-state) in the cryoEM data set because the occupancy of Rubisco was not 100%. We would expect cryoEM reconstructions of substrate-free GroEL ADP·BeF_x to have a *trans* ring in the T-state. However, our data did not support this. In addition, the negative cooperativity exerted by Rubisco binding is thought to be present in the binary complex, independent of ATP and GroES.

Additional studies are needed to properly uncover the functional significance of this *trans* ring conformation.

8.7. Conclusion

The work reported in this thesis provides a novel insight into protein folding by the GroEL-GroES system. By following the trajectory of a stringent substrate through the chaperonin reaction cycle, we have uncovered novel molecular mechanisms underlying intermediate steps. Binding of non-native substrate to GroEL is classically thought to be mediated by hydrophobic residues. Our 4.5 Å cryoEM structure of GroEL-Rubisco highlighted the importance of non-classical substrate binding residues, such as Met-267 and Arg-268. This provides additional avenues for further study. What is the relative importance of the charged residue, Arg-268, in maintaining substrate contacts? Are these contacts specific to a certain substrate or certain type of substrate? Or are they involved in binding all GroEL substrates? Is Met-267 involved in transmitting an allosteric signal to the *trans* ring, thereby inhibiting further substrate binding? The 3.4 Å cryoEM structure of GroEL-Rubisco-ADP·BeF_x answers a previously unresolved question; How does GroEL recruit GroES while preventing escape of non-native substrate? Our structure answers this by assigning two roles to the GroEL subunits. Four subunits maintain contact with bound substrate, and the other three detach and recruit GroES. While the interaction between non-native substrate and GroEL apical domains may be weaker in this intermediate state, an interaction with all seven GroEL C-terminal tails might offset this. Our structure also hints at a mechanism for the forced unfolding of substrate upon binding of GroES. Do alternate substrates yield different intermediate conformations of GroEL? Or is this a common GroEL intermediate. Lastly, our structure of GroEL-GroES-Rubisco-ADP·AlF_x showed native (or near-native) Rubisco in the *cis* folding chamber. Interactions between the encapsulated substrate and the GroEL-GroES cavity hint at possible roles in folding for several GroEL residues. Would substrates still fold inside GroEL-GroES if these residues were mutated?

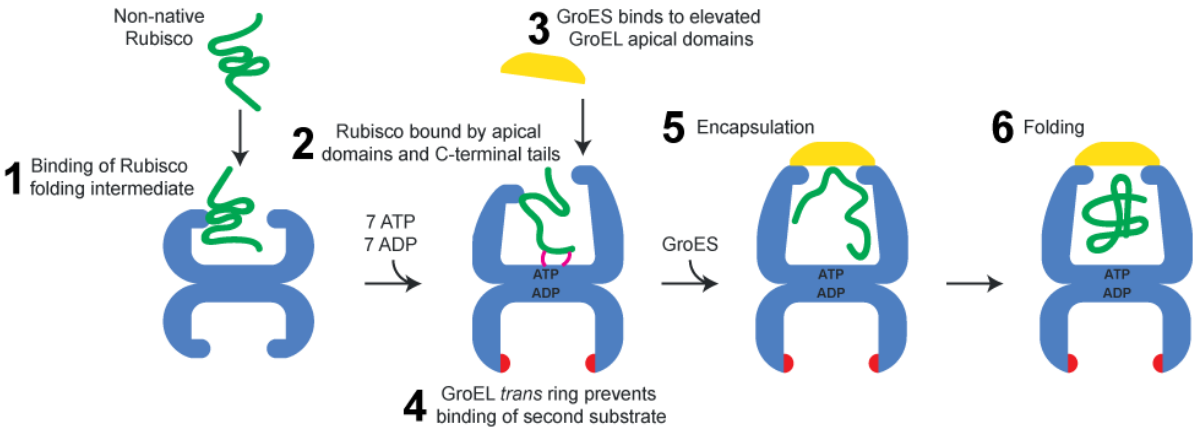

Figure 8-1. Proposed model for binding, encapsulation and folding of Rubisco by GroEL-GroES. (**1**) A Rubisco folding intermediate binds to consecutive GroEL subunits. (**2**) Upon binding of ATP to the cis ring, three GroEL subunits elevate. Non-native substrate is sequestered by the other four GroEL subunits, and the collective C-terminal tails of cis ring subunits. (**3**) GroES binds to the elevated GroEL apical domains. (**4**) A conformational change in the *trans* ring prevents binding of a second substrate, perhaps until ATP as hydrolysed in the *cis* ring. (**5**) Upon GroES binding, non-native Rubisco is encapsulated. (**6**) The Rubisco monomer folds to a native or near-native state.

9. References

Anfinsen, C., 1973. Principles that Govern the Folding of Protein Chains. *Science*, 181(4096), pp.223-230.

Anfinsen, C., Haber, E., Sela, M. and White, F., 1961. Kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences*, 47(9), pp.1309-1314.

Apetri, A. and Horwich, A., 2008. Chaperonin chamber accelerates protein folding through passive action of preventing aggregation. *Proceedings of the National Academy of Sciences*, 105(45), pp.17351-17355.

Azia, A., Unger, R. and Horovitz, A., 2012. What distinguishes GroEL substrates from other Escherichia coli proteins?. *FEBS Journal*, 279(4), pp.543-550.

Bai, X., McMullan, G. and Scheres, S., 2015. How cryoEM is revolutionizing structural biology. *Trends in Biochemical Sciences*, 40(1), pp.49-57.

Baker, L. and Rubinstein, J., 2010. Radiation Damage in Electron Cryomicroscopy. *Methods in Enzymology*, pp.371-388.

Balchin, D., Hayer-Hartl, M. and Hartl, F., 2016. In vivo aspects of protein folding and quality control. *Science*, 353(6294), pp.aac4354-aac4354.

Balchin, D., Miličić, G., Strauss, M., Hayer-Hartl, M. and Hartl, F., 2018. Pathway of Actin Folding Directed by the Eukaryotic Chaperonin TRiC. *Cell*, 174(6), pp.1507-1521.e16.

Baldwin, R., 1994. Matching speed and stability. *Nature*, 369(6477), pp.183-184.

Bhushan, S., Gartmann, M., Halic, M., Armache, J., Jarasch, A., Mielke, T., Berninghausen, O., Wilson, D. and Beckmann, R., 2010. α-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nature Structural & Molecular Biology*, 17(3), pp.313-317.

Bogumil, D. and Dagan, T., 2012. Cumulative Impact of Chaperone-Mediated Folding on Genome Evolution. *Biochemistry*, 51(50), pp.9941-9953.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D., Joachimiak, A., Horwich, A. and Sigler, P., 1994. The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature*, 371(6498), pp.578-586.

Brandts, J., Halvorson, H. and Brennan, M., 1975. Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry*, 14(22), pp.4953-4963.

Brinker, A., Pfeifer, G., Kerner, M., Naylor, D., Hartl, F. and Hayer-Hartl, M., 2001. Dual Function of Protein Confinement in Chaperonin-Assisted Protein Folding. *Cell*, 107(2), pp.223-233.

Bryngelson, J. and Wolynes, P., 1987. Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of Sciences*, 84(21), pp.7524-7528.

Cavalli, A., Haberthür, U., Paci, E. and Caflisch, A., 2003. Fast protein folding on downhill energy landscape. *Protein Science*, 12(8), pp.1801-1803.

Cayley, S., Lewis, B., Guttman, H. and Record, M., 1991. Characterization of the cytoplasm of Escherichia coli K-12 as a function of external osmolarity. *Journal of Molecular Biology*, 222(2), pp.281-300.

Chakraborty, K., Chatila, M., Sinha, J., Shi, Q., Poschner, B., Sikor, M., Jiang, G., Lamb, D., Hartl, F. and Hayer-Hartl, M., 2010. Chaperonin-Catalyzed Rescue of Kinetically Trapped States in Protein Folding. *Cell*, 142(1), pp.112-122.

Chaudhry, C., Farr, G., Todd, M., Rye, H., Brunger, A., Adams, P., Horwich, A. and Sigler, P., 2003. Role of the γ-phosphate of ATP in triggering protein folding by GroEL-GroES: function, structure and energetics. *The EMBO Journal*, 22(19), pp.4877-4887.

Chaudhuri, T., Farr, G., Fenton, W., Rospert, S. and Horwich, A., 2001. GroEL/GroES-Mediated Folding of a Protein Too Large to Be Encapsulated. *Cell*, 107(2), pp.235-246.

Chen, D., Madan, D., Weaver, J., Lin, Z., Schröder, G., Chiu, W. and Rye, H., 2013. Visualizing GroEL/ES in the Act of Encapsulating a Folding Protein. *Cell*, 153(6), pp.1354-1365.

Chen, J., Noble, A., Kang, J. and Darst, S., 2019. Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO. *Journal of Structural Biology: X*, 1, p.100005.

Chen, S., Roseman, A., Hunter, A., Wood, S., Burston, S., Ranson, N., Clarke, A. and Saibil, H., 1994. Location of a folding protein and shape changes in GroEL–GroES complexes imaged by cryo-electron microscopy. *Nature*, 371(6494), pp.261-264.

Cheng, Y., Grigorieff, N., Penczek, P. and Walz, T., 2015. A Primer to Single-Particle Cryo-Electron Microscopy. *Cell*, 161(3), pp.438-449.

Cheng, K., Wilkinson, M., Chaban, Y. and Wigley, D., 2020. A conformational switch in response to Chi converts RecBCD from phage destruction to DNA repair. *Nature Structural & Molecular Biology*, 27(1), pp.71-77.

Clare, D., Bakkes, P., van Heerikhuizen, H., van der Vies, S. and Saibil, H., 2009. Chaperonin complex with a newly folded protein encapsulated in the folding chamber. *Nature*, 457(7225), pp.107-110.

Clare, D., Vasishtan, D., Stagg, S., Quispe, J., Farr, G., Topf, M., Horwich, A. and Saibil, H., 2012. ATP-Triggered Conformational Changes Delineate Substrate-Binding and -Folding Mechanics of the GroEL Chaperonin. *Cell*, 149(1), pp.113-123.

Clare, D., Siebert, C., Hecksel, C., Hagen, C., Mordhorst, V., Grange, M., Ashton, A., Walsh, M., Grünewald, K., Saibil, H., Stuart, D. and Zhang, P., 2017. Electron Bio-Imaging Centre (eBIC):

the UK national research facility for biological electron microscopy. *Acta Crystallographica Section D Structural Biology*, 73(6), pp.488-495.

Craig, E., Eisenman, H. and Hundley, H., 2003. Ribosome-tethered molecular chaperones: the first line of defense against protein misfolding?. *Current Opinion in Microbiology*, 6(2), pp.157-162.

Dahiya, V. and Chaudhuri, T., 2013. Chaperones GroEL/GroES Accelerate the Refolding of a Multidomain Protein through Modulating On-pathway Intermediates. *Journal of Biological Chemistry*, 289(1), pp.286-298.

Dandey, V., Wei, H., Zhang, Z., Tan, Y., Acharya, P., Eng, E., Rice, W., Kahn, P., Potter, C. and Carragher, B., 2018. Spotiton: New features and applications. *Journal of Structural Biology*, 202(2), pp.161-169.

Deller, M., Kong, L. and Rupp, B., 2016. Protein stability: a crystallographer's perspective. *Acta Crystallographica Section F Structural Biology Communications*, 72(2), pp.72-95.

DePristo, M., Weinreich, D. and Hartl, D., 2005. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nature Reviews Genetics*, 6(9), pp.678-687.

D'Imprima, E., Floris, D., Joppe, M., Sánchez, R., Grininger, M. and Kühlbrandt, W., 2019. Protein denaturation at the air-water interface and how to prevent it. *eLife*, 8.

D'Imprima, E. and Kühlbrandt, W., 2021. Current limitations to high-resolution structure determination by single-particle cryoEM. *Quarterly Reviews of Biophysics*, 54.

Dobson, C., 2003. Protein folding and misfolding. *Nature*, 426(6968), pp.884-890.

Dobson, C., Šali, A. and Karplus, M., 1998. Protein Folding: A Perspective from Theory and Experiment. *Angewandte Chemie International Edition*, 37(7), pp.868-893.

Dubochet, J., Adrian, M., Chang, J., Homo, J., Lepault, J., McDowall, A. and Schultz, P., 1988. Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics*, 21(2), pp.129-228.

Dyachenko, A., Gruber, R., Shimon, L., Horovitz, A. and Sharon, M., 2013. Allosteric mechanisms can be distinguished using structural mass spectrometry. *Proceedings of the National Academy of Sciences*, 110(18), pp.7235-7239.

Elad, N., Farr, G., Clare, D., Orlova, E., Horwich, A. and Saibil, H., 2007. Topologies of a Substrate Protein Bound to the Chaperonin GroEL. *Molecular Cell*, 26(3), pp.415-426.

Ellis, R. and Minton, A., 2006. Protein aggregation in crowded environments. *Biological Chemistry*, 387(5).

Ellis, R., 1996. Discovery of molecular chaperones. *Cell Stress & Chaperones*, 1(3), pp.155-160.

EMBL-EBI, 2021. *PDBe*. [online] EMBL-EBI. Available from: https://www.ebi.ac.uk/pdbe/ [Accessed 1 Mar. 2021].

Englander, S. and Mayne, L., 2017. The case for defined protein folding pathways. *Proceedings of the National Academy of Sciences*, 114(31), pp.8253-8258.

Fares, M., Ruiz-González, M., Moya, A., Elena, S. and Barrio, E., 2002. Endosymbiotic bacteria: GroEL buffers against deleterious mutations. *Nature*, 417(6887), pp.398-398.

Farr, G., Fenton, W., Chaudhuri, T., Clare, D., Saibil, H. and Horwich, A., 2003. Folding with and without encapsulation by cis- and trans-only GroEL-GroES complexes. *The EMBO Journal*, 22(13), pp.3220-3230.

Farr, G., Furtak, K., Rowland, M., Ranson, N., Saibil, H., Kirchhausen, T. and Horwich, A., 2000. Multivalent Binding of Nonnative Substrate Proteins by the Chaperonin GroEL. *Cell*, 100(5), pp.561-573.

Fayet, O., Ziegelhoffer, T. and Georgopoulos, C., 1989. The groES and groEL heat shock gene products of Escherichia coli are essential for bacterial growth at all temperatures. *Journal of Bacteriology*, 171(3), pp.1379-1385.

Fenton, W. and Horwich, A., 1997. GroEL-Mediated protein folding. *Protein Science*, 6(4), pp.743-760.

Fenton, W., Kashi, Y., Furtak, K. and Horwich, A., 1994. Residues in chaperonin GroEL required for polypeptide binding and release. *Nature*, 371(6498), pp.614-619.

Frueh, D., Goodrich, A., Mishra, S. and Nichols, S., 2013. NMR methods for structural studies of large monomeric and multimeric proteins. *Current Opinion in Structural Biology*, 23(5), pp.734-739.

Frydman, J., Erdjument-Bromage, H., Tempst, P. and Hartl, F., 1999. *Nature Structural Biology*, 6(7), pp.697-705.

Fujiwara, K., Ishihama, Y., Nakahigashi, K., Soga, T. and Taguchi, H., 2010. A systematic survey of in vivo obligate chaperonin-dependent substrates. *The EMBO Journal*, 29(9), pp.1552-1564.

Georgescauld, F., Popova, K., Gupta, A., Bracher, A., Engen, J., Hayer-Hartl, M. and Hartl, F., 2014. GroEL/ES Chaperonin Modulates the Mechanism and Accelerates the Rate of TIM-Barrel Domain Folding. *Cell*, 157(4), pp.922-934.

Georgopoulos, C., 1992. The emergence of the chaperone machines. *Trends in Biochemical Sciences*, 17(8), pp.295-299.

Georgopoulos, C., Hendrix, R., Casjens, S. and Kaiser, A., 1973. Host participation in bacteriophage lambda head assembly. *Journal of Molecular Biology*, 76(1), pp.45-60.

Glaeser, R., 1971. Limitations to significant information in biological electron microscopy as a result of radiation damage. *Journal of Ultrastructure Research*, 36(3-4), pp.466-482.

Glaeser, R., 2016. How good can cryoEM become? *Nature Methods*, 13(1), pp.28-32.

Glaeser, R., 2019. How Good Can Single-Particle CryoEM Become? What Remains Before It Approaches Its Physical Limits? *Annual Review of Biophysics*, 48(1), pp.45-61.

Goldberg, M., Semisotnov, G., Friguet, B., Kuwajima, K., Ptitsyn, O. and Sugai, S., 1990. An early immunoreactive folding intermediate of the tryptophan synthase β2 subunit is a 'molten globule'. *FEBS Letters*, 263(1), pp.51-56.

Goloubinoff, P., Christeller, J., Gatenby, A. and Lorimer, G., 1989. Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature*, 342(6252), pp.884-889.

Gupta, A., Haldar, S., Miličić, G., Hartl, F. and Hayer-Hartl, M., 2014. Active Cage Mechanism of Chaperonin-Assisted Protein Folding Demonstrated at Single-Molecule Level. *Journal of Molecular Biology*, 426(15), pp.2739-2754.

Hartl, F., 1996. Molecular chaperones in cellular protein folding. *Nature*, 381(6583), pp.571-580.

Hayer-Hartl, M., Bracher, A. and Hartl, F., 2016. The GroEL–GroES Chaperonin Machine: A Nano-Cage for Protein Folding. *Trends in Biochemical Sciences*, 41(1), pp.62-76.

Hemmingsen, S., Woolford, C., van der Vies, S., Tilly, K., Dennis, D., Georgopoulos, C., Hendrix, R. and Ellis, R., 1988. Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature*, 333(6171), pp.330-334.

Hendrix, R., 1979. Purification and properties of groE, a host protein involved in bacteriophage assembly. *Journal of Molecular Biology*, 129(3), pp.375-392.

Herzik, M., Wu, M. and Lander, G., 2019. High-resolution structure determination of sub-100 kDa complexes using conventional cryoEM. *Nature Communications*, 10(1).

Hohn, T., Hohn, B., Engel, A., Wurtz, M. and Smith, P., 1979. Isolation and characterization of the host protein groE involved in bacteriophage lambda assembly. *Journal of Molecular Biology*, 129(3), pp.359-373.

Horst, R., Fenton, W., Englander, S., Wuthrich, K. and Horwich, A., 2007. Folding trajectories of human dihydrofolate reductase inside the GroEL-GroES chaperonin cavity and free in solution. *Proceedings of the National Academy of Sciences*, 104(52), pp.20788-20792.

Houry, W., Frishman, D., Eckerskorn, C., Lottspeich, F. and Hartl, F., 1999. Identification of in vivo substrates of the chaperonin GroEL. *Nature*, 402(6758), pp.147-154.

Hunt, J., Weaver, A., Landry, S., Gierasch, L. and Deisenhofer, J., 1996. The crystal structure of the GroES co-chaperonin at 2.8 Å resolution. *Nature*, 379(6560), pp.37-45.

Ishimoto, T., Fujiwara, K., Niwa, T. and Taguchi, H., 2014. Conversion of a Chaperonin GroEL-independent Protein into an Obligate Substrate. *Journal of Biological Chemistry*, 289(46), pp.32073-32080.

Jackson, S. and Fersht, A., 1991. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry*, 30(43), pp.10428-10435.

Jewett, A. and Shea, J., 2009. Reconciling theories of chaperonin accelerated folding with experimental evidence. *Cellular and Molecular Life Sciences*, 67(2), pp.255-276.

Kaiser, C., Goldman, D., Chodera, J., Tinoco, I. and Bustamante, C., 2011. The Ribosome Modulates Nascent Protein Folding. *Science*, 334(6063), pp.1723-1727.

Karplus, M. and Weaver, D., 1976. Protein-folding dynamics. *Nature*, 260(5550), pp.404-406.

Karplus, M., 1997. The Levinthal paradox: yesterday and today. *Folding and Design*, 2, pp.S69-S75.

Karplus, M., 2011. Behind the folding funnel diagram. *Nature Chemical Biology*, 7(7), pp.401-404.

Kendrew, J., Bobo, G., Dintzis, H., Parrish, R., Wyckoff, H. and Phillips, D., 1958. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature*, 181(4610), pp.662-666.

Kerner, M., Naylor, D., Ishihama, Y., Maier, T., Chang, H., Stines, A., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M. and Hartl, F., 2005. Proteome-wide Analysis of Chaperonin-Dependent Protein Folding in Escherichia coli. *Cell*, 122(2), pp.209-220.

Kim, P. and Baldwin, R., 1990. Intermediates in the Folding Reactions of Small Proteins. *Annual Review of Biochemistry*, 59(1), pp.631-660.

Kim, Y., Hipp, M., Bracher, A., Hayer-Hartl, M. and Ulrich Hartl, F., 2013. Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annual Review of Biochemistry*, 82(1), pp.323-355.

Klebl, D., Gravett, M., Kontziampasis, D., Wright, D., Bon, R., Monteiro, D., Trebbin, M., Sobott, F., White, H., Darrow, M., Thompson, R. and Muench, S., 2020. Need for Speed: Examining Protein Behaviour during CryoEM Grid Preparation at Different Timescales. *Structure*, 28(11), pp.1238-1248.e4.

Klebl, D., Feasey, M., Hesketh, E., Ranson, N., Wurdak, H., Sobott, F., Bon, R. and Muench, S., 2021. CryoEM structure of human mitochondrial HSPD1. *iScience*, 24(1), p.102022.

Kontziampasis, D., Klebl, D., Iadanza, M., Scarff, C., Kopf, F., Sobott, F., Monteiro, D., Trebbin, M., Muench, S. and White, H., 2019. A cryo-EM grid preparation device for time-resolved structural studies. IUCrJ, 6(6), pp.1024-1031.

Kuhlbrandt, W., 2014. The Resolution Revolution. *Science*, 343(6178), pp.1443-1444.

Langer, T., Pfeifer, G., Martin, J., Baumeister, W. and Hartl, F., 1992. Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein substrate within its central cavity. *The EMBO Journal*, 11(13), pp.4757–4765.

Levinthal, C., 1968. Are there pathways for protein folding? *Journal de Chimie Physique*, 65, pp.44-45.

Levinthal, C., 1969. How to fold graciously. *Mössbaun Spectroscopy in Biological Systems Proceedings*, 67(41), pp.22–24.

Liebschner, D., Afonine, P., Baker, M., Bunkóczi, G., Chen, V., Croll, T., Hintze, B., Hung, L., Jain, S., McCoy, A., Moriarty, N., Oeffner, R., Poon, B., Prisant, M., Read, R., Richardson, J., Richardson, D., Sammito, M., Sobolev, O., Stockwell, D., Terwilliger, T., Urzhumtsev, A., Videau, L., Williams, C. and Adams, P., 2019. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D Structural Biology*, 75(10), pp.861-877.

Lim, N. and Jackson, S., 2015. Mechanistic Insights into the Folding of Knotted Proteins In Vitro and In Vivo. *Journal of Molecular Biology*, 427(2), pp.248-258.

Lin, Z. and Rye, H., 2004. Expansion and Compression of a Protein Folding Intermediate by GroEL. *Molecular Cell*, 16(1), pp.23-34.

Lin, Z., Madan, D. and Rye, H., 2008. GroEL stimulates protein folding through forced unfolding. *Nature Structural & Molecular Biology*, 15(3), pp.303-311.

Lin, Z., Puchalla, J., Shoup, D. and Rye, H., 2013. Repetitive Protein Unfolding by the trans ring of the GroEL-GroES Chaperonin Complex Stimulates Folding. *Journal of Biological Chemistry*, 288(43), pp.30944-30955.

Lindorff-Larsen, K., Piana, S., Dror, R. and Shaw, D., 2011. How Fast-Folding Proteins Fold. *Science*, 334(6055), pp.517-520.

Lopez, T., Dalton, K. and Frydman, J., 2015. The Mechanism and Function of Group II Chaperonins. *Journal of Molecular Biology*, 427(18), pp.2919-2930.

Machida, K., Kono-Okada, A., Hongo, K., Mizobata, T. and Kawata, Y., 2008. Hydrophilic Residues526KNDAAD531in the Flexible C-terminal Region of the Chaperonin GroEL Are

Critical for Substrate Protein Folding within the Central Cavity. *Journal of Biological Chemistry*, 283(11), pp.6886-6896.

Mayer, M., 2013. Hsp70 chaperone dynamics and molecular mechanism. *Trends in Biochemical Sciences*, 38(10), pp.507-514.

Mayhew, M., da Silva, A., Martin, J., Erdjument-Bromage, H., Tempst, P. and Hartl, F., 1996. Protein folding in the central cavity of the GroEL–GroES chaperonin complex. *Nature*, 379(6564), pp.420-426.

McMullan, G., Faruqi, A. and Henderson, R., 2016. Direct Electron Detectors. *Methods in Enzymology*, pp.1-17.

McMullin, T. and Hallberg, R., 1987. A normal mitochondrial protein is selectively synthesized and accumulated during heat shock in Tetrahymena thermophila. *Molecular and Cellular Biology*, 7(12), pp.4414-4423.

McMullin, T. and Hallberg, R., 1988. A highly evolutionarily conserved mitochondrial protein is structurally related to the protein encoded by the Escherichia coli groEL gene. *Molecular and Cellular Biology*, 8(1), pp.371-380.

Meric, G., Robinson, A. and Roberts, C., 2017. Driving Forces for Nonnative Protein Aggregation and Approaches to Predict Aggregation-Prone Regions. *Annual Review of Chemical and Biomolecular Engineering*, 8(1), pp.139-159.

Mueller-Cajar, O. and Whitney, S., 2008. Evolving improved Synechococcus Rubisco functional expression in Escherichia coli. *Biochemical Journal*, 414(2), pp.205-214.

Murata, K. and Wolf, M., 2018. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1862(2), pp.324-334.

Naydenova, K., Peet, M. and Russo, C., 2019. Multifunctional graphene supports for electron cryomicroscopy. *Proceedings of the National Academy of Sciences*, p.201904766.

Nilsson, O., Hedman, R., Marino, J., Wickles, S., Bischoff, L., Johansson, M., Müller-Lucks, A., Trovato, F., Puglisi, J., O'Brien, E., Beckmann, R. and von Heijne, G., 2015. Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Reports*, 12(10), pp.1533-1540.

Nobel Media AB, 2021. *The Nobel Prize in Chemistry 1972*. [online] Nobelprize.org. Available from: https://www.nobelprize.org/nobel_prizes/chemistry/laureates/1972/ [Accessed 1 Mar. 2021].

Noble, A., Dandey, V., Wei, H., Brasch, J., Chase, J., Acharya, P., Tan, Y., Zhang, Z., Kim, L., Scapin, G., Rapp, M., Eng, E., Rice, W., Cheng, A., Negro, C., Shapiro, L., Kwong, P., Jeruzalmi, D., des Georges, A., Potter, C. and Carragher, B., 2018. Routine single particle CryoEM sample and grid characterization by tomography. *eLife*, 7.

Nogales, E., 2015. The development of cryoEM into a mainstream structural biology technique. *Nature Methods*, 13(1), pp.24-27.

Orlova, E. and Saibil, H., 2011. Structural Analysis of Macromolecular Assemblies by Electron Microscopy. *Chemical Reviews*, 111(12), pp.7710-7748.

Ostermann, J., Horwich, A., Neupert, W. and Hartl, F., 1989. Protein folding in mitochondria requires complex formation with hsp60 and ATP hydrolysis. *Nature*, 341(6238), pp.125-130.

Ott, A., Locher, L., Koch, M. and Deuerling, E., 2015. Functional Dissection of the Nascent Polypeptide-Associated Complex in Saccharomyces cerevisiae. *PLOS ONE*, 10(11), p.e0143457.

Passmore, L. and Russo, C., 2016. Specimen Preparation for High-Resolution CryoEM. *Methods in Enzymology*, pp.51-86.

Paul, S., Singh, C., Mishra, S. and Chaudhuri, T., 2007. The 69 kDa Escherichia coli maltodextrin glucosidase does not get encapsulated underneath GroES and folds through trans mechanism during GroEL/GroES-assisted folding. *The FASEB Journal*, 21(11), pp.2874-2885.

Pelham, H., 1986. Speculations on the functions of the major heat shock and glucose-regulated proteins. *Cell*, 46(7), pp.959-961.

Penczek, P., 2010. Resolution Measures in Molecular Electron Microscopy. *Methods in Enzymology*, pp.73-100.

Pettersen, E., Goddard, T., Huang, C., Couch, G., Greenblatt, D., Meng, E. and Ferrin, T., 2004. UCSF Chimera: A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), pp.1605-1612.

Pettersen, E., Goddard, T., Huang, C., Meng, E., Couch, G., Croll, T., Morris, J. and Ferrin, T., 2020. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1), pp.70-82.

Plotkin, S. and Onuchic, J., 2002. Understanding protein folding with energy landscape theory Part I: Basic concepts. *Quarterly Reviews of Biophysics*, 35(02), pp.111–167.

Priya, S., Sharma, S., Sood, V., Mattoo, R., Finka, A., Azem, A., De Los Rios, P. and Goloubinoff, P., 2013. GroEL and CCT are catalytic unfoldases mediating out-of-cage polypeptide refolding without ATP. *Proceedings of the National Academy of Sciences*, 110(18), pp.7199-7204.

Pushkin, A., Tsuprun, V., Solovjeva, N., Shubin, V., Evstigneeva, Z. and Kretovich, W., 1982. High molecular weight pea leaf protein similar to the groE protein of Escherichia coli. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 704(2), pp.379-384.

Qiu, L., Pabit, S., Roitberg, A. and Hagen, S., 2002. Smaller and Faster: The 20-Residue Trp-Cage Protein Folds in 4 μs. *Journal of the American Chemical Society*, 124(44), pp.12952-12953.

Radivojac, P., 2004. Protein flexibility and intrinsic disorder. *Protein Science*, 13(1), pp.71-80.

Ranson, N., Farr, G., Roseman, A., Gowen, B., Fenton, W., Horwich, A. and Saibil, H., 2001. ATP-Bound States of GroEL Captured by Cryo-Electron Microscopy. *Cell*, 107(7), pp.869-879.

RCSB, 2021. *RCSB PDB - PDB Statistics*. [online] RCSB. Available from: http://www.rcsb.org/pdb/stats [Accessed 1 Mar. 2021].

Roseman, A., Chen, S., White, H., Braig, K. and Saibil, H., 1996. The Chaperonin ATPase Cycle: Mechanism of Allosteric Switching and Movements of Substrate-Binding Domains in GroEL. *Cell*, 87(2), pp.241-251.

Rosenthal, P. and Henderson, R., 2003. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *Journal of Molecular Biology*, 333(4), pp.721-745.

Rudiger, S., 1997. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *The EMBO Journal*, 16(7), pp.1501-1507.

Russo, C. and Passmore, L., 2014. Ultrastable gold substrates for electron cryomicroscopy. *Science*, 346(6215), pp.1377-1380.

Rutherford, S. and Lindquist, S., 1998. Hsp90 as a capacitor for morphological evolution. *Nature*, 396(6709), pp.336-342.

Rye, H., Roseman, A., Chen, S., Furtak, K., Fenton, W., Saibil, H. and Horwich, A., 1999. GroEL-GroES Cycling. *Cell*, 97(3), pp.325-338.

Saibil, H., 2013. Chaperone machines for protein folding, unfolding and disaggregation. *Nature Reviews Molecular Cell Biology*, 14(10), pp.630-642.

Saibil, H., Fenton, W., Clare, D. and Horwich, A., 2013. Structure and Allostery of the Chaperonin GroEL. *Journal of Molecular Biology*, 425(9), pp.1476-1487.

Saibil, H., Zheng, D., Roseman, A., Hunter, A., Watson, G., Chen, S., auf der Mauer, A., O'Hara, B., Wood, S., Mann, N., Barnett, L. and Ellis, R., 1993. ATP induces large quaternary rearrangements in a cage-like chaperonin structure. *Current Biology*, 3(5), pp.265-273.

Sali, A., Shakhnovich, E. and Karplus, M., 1994. How does a protein fold? *Nature*, 369(6477), pp.248-251.

Scheres, S. and Chen, S., 2012. Prevention of overfitting in cryoEM structure determination. *Nature Methods*, 9(9), pp.853-854.

Scheres, S., 2012a. A Bayesian View on CryoEM Structure Determination. *Journal of Molecular Biology*, 415(2), pp.406-418.

Scheres, S., 2012b. RELION: Implementation of a Bayesian approach to cryoEM structure determination. *Journal of Structural Biology*, 180(3), pp.519-530.

Scheres, S., 2016. Processing of Structurally Heterogeneous CryoEM Data in RELION. *Methods in Enzymology*, pp.125-157.

Schütz, S. and Sprangers, R., 2020. Methyl TROSY spectroscopy: A versatile NMR approach to study challenging biological systems. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 116, pp.56-84.

Shtilerman, M., Lorimer, G. and Englander, S., 1999. Chaperonin Function: Folding by Forced Unfolding. *Science*, 284(5415), pp.822-825.

Sigler, P., Xu, Z., Rye, H., Burston, S., Fenton, W. and Horwich, A., 1998. Structure and Function in GroEL-Mediated Protein Folding. *Annual Review of Biochemistry*, 67(1), pp.581-608.

Sigworth, F., 2015. Principles of cryoEM single-particle image processing. *Microscopy*, 65(1), pp.57-67.

Sivalingam, G., Yan, J., Sahota, H. and Thalassinos, K., 2013. Amphitrite: A program for processing travelling wave ion mobility mass spectrometry data. *International Journal of Mass Spectrometry*, 345-347, pp.54-62.

Soler, M. and Faísca, P., 2013. Effects of Knots on Protein Folding Properties. *PLoS ONE*, 8(9), p.e74755.

Stan, G., Lorimer, G., Thirumalai, D. and Brooks, B., 2007. Coupling between allosteric transitions in GroEL and assisted folding of a substrate protein. *Proceedings of the National Academy of Sciences*, 104(21), pp.8803-8808.

Swain, J. and Gierasch, L., 2005. First glimpses of a chaperonin-bound folding intermediate. *Proceedings of the National Academy of Sciences*, 102(39), pp.13715-13716.

Tan, Y., Baldwin, P., Davis, J., Williamson, J., Potter, C., Carragher, B. and Lyumkis, D., 2017. Addressing preferred specimen orientation in single-particle cryoEM through tilting. *Nature Methods*, 14(8), pp.793-796.

Tang, G., Peng, L., Baldwin, P., Mann, D., Jiang, W., Rees, I. and Ludtke, S., 2007. EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*, 157(1), pp.38-46.

Tang, Y., Chang, H., Chakraborty, K., Hartl, F. and Hayer-Hartl, M., 2008. Essential role of the chaperonin folding compartment in vivo. *The EMBO Journal*.

Tang, Y., Chang, H., Roeben, A., Wischnewski, D., Wischnewski, N., Kerner, M., Hartl, F. and Hayer-Hartl, M., 2006. Structural Features of the GroEL-GroES Nano-Cage Required for Rapid Folding of Encapsulated Protein. *Cell*, 125(5), pp.903-914.

Thommen, M., Holtkamp, W. and Rodnina, M., 2017. Co-translational protein folding: progress and methods. *Current Opinion in Structural Biology*, 42, pp.83-89.

Todd, M., Lorimer, G. and Thirumalai, D., 1996. Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism. *Proceedings of the National Academy of Sciences*, 93(9), pp.4030-4035.

Todd, M., Viitanen, P. and Lorimer, G., 1994. Dynamics of the chaperonin ATPase cycle: implications for facilitated protein folding. *Science*, 265(5172), pp.659-666.

Tokuriki, N. and Tawfik, D., 2009a. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature*, 459(7247), pp.668-673.

Tokuriki, N. and Tawfik, D., 2009b. Stability effects of mutations and protein evolvability. *Current Opinion in Structural Biology*, 19(5), pp.596-604.

Tokuriki, N., Stricher, F., Serrano, L. and Tawfik, D., 2008. How Protein Stability and New Functions Trade Off. *PLoS Computational Biology*, 4(2), p.e1000002.

Trent, J., Nimmesgern, E., Wall, J., Hartl, F. and Horwich, A., 1991. A molecular chaperone from a thermophilic archaebacterium is related to the eukaryotic protein t-complex polypeptide-1. *Nature*, 354(6353), pp.490-493.

Tyagi, N., Fenton, W., Deniz, A. and Horwich, A., 2011. Double mutant MBP refolds at same rate in free solution as inside the GroEL/GroES chaperonin chamber when aggregation in free solution is prevented. *FEBS Letters*, 585(12), pp.1969-1972.

van Duijn, E., Simmons, D., van den Heuvel, R., Bakkes, P., van Heerikhuizen, H., Heeren, R., Robinson, C., van der Vies, S. and Heck, A., 2006. Tandem Mass Spectrometry of Intact GroEL−Substrate Complexes Reveals Substrate-Specific Conformational Changes in the trans Ring. *Journal of the American Chemical Society*, 128(14), pp.4694-4702.

van Duijn, E., Heck, A. and van der Vies, S., 2007. Inter-ring communication allows the GroEL chaperonin complex to distinguish between different substrates. *Protein Science*, 16(5), pp.956-965.

Verba, K., Wang, R., Arakawa, A., Liu, Y., Shirouzu, M., Yokoyama, S. and Agard, D., 2016. Atomic structure of Hsp90-Cdc37-Cdk4 reveals that Hsp90 traps and stabilizes an unfolded kinase. *Science*, 352(6293), pp.1542-1547.

Wang, J., Herman, C., Tipton, K., Gross, C. and Weissman, J., 2002. Directed Evolution of Substrate-Optimized GroEL/S Chaperonins. *Cell*, 111(7), pp.1027-1039.

Wang, J. and Chen, L., 2003. Domain Motions in GroEL upon Binding of an Oligopeptide. *Journal of Molecular Biology*, 334(3), pp.489-499.

Weaver, J. and Rye, H., 2014. The C-terminal Tails of the Bacterial Chaperonin GroEL Stimulate Protein Folding by Directly Altering the Conformation of a Substrate Protein. *Journal of Biological Chemistry*, 289(33), pp.23219-23232.

Weaver, J., Jiang, M., Roth, A., Puchalla, J., Zhang, J. and Rye, H., 2017. GroEL actively stimulates folding of the endogenous substrate protein PepQ. *Nature Communications*, 8, p.15934.

Wei, H., Dandey, V., Zhang, Z., Raczkowski, A., Rice, W., Carragher, B. and Potter, C., 2018. Optimizing "self-wicking" nanowire grids. *Journal of Structural Biology*, 202(2), pp.170-174.

Weissman, J., Rye, H., Fenton, W., Beechem, J. and Horwich, A., 1996. Characterization of the Active Intermediate of a GroEL–GroES-Mediated Protein Folding Reaction. *Cell*, 84(3), pp.481-490.

Wetlaufer, D., 1973. Nucleation, Rapid Folding, and Globular Intrachain Regions in Proteins. *Proceedings of the National Academy of Sciences*, 70(3), pp.697-701.

Williams, C., Headd, J., Moriarty, N., Prisant, M., Videau, L., Deis, L., Verma, V., Keedy, D., Hintze, B., Chen, V., Jain, S., Lewis, S., Arendall, W., Snoeyink, J., Adams, P., Lovell, S., Richardson, J. and Richardson, D., 2017. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science*, 27(1), pp.293-315.

Williams, T. and Fares, M., 2010. The Effect of Chaperonin Buffering on Protein Evolution. *Genome Biology and Evolution*, 2(0), pp.609-619.

Wolynes, P., Onuchic, J. and Thirumalai, D., 1995. Navigating the folding routes. *Science*, 267(5204), pp.1619-1620.

Xu, Z., Horwich, A. and Sigler, P., 1997. The crystal structure of the asymmetric GroEL–GroES–(ADP)7 chaperonin complex. *Nature*, 388(6644), pp.741-750.

Zhang, K., 2016. Gctf: Real-time CTF determination and correction. *Journal of Structural Biology*, 193(1), pp.1-12.

Zimmerman, S. and Trach, S., 1991. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *Journal of Molecular Biology*, 222(3), pp.599-620.

Zivanov, J., Nakane, T. and Scheres, S., 2019. A Bayesian approach to beam-induced motion correction in cryoEM single-particle analysis. *IUCrJ*, 6(1), pp.5-17.

Zivanov, J., Nakane, T., Forsberg, B., Kimanius, D., Hagen, W., Lindahl, E. and Scheres, S., 2018. New tools for automated high-resolution cryoEM structure determination in RELION-3. *eLife*, 7.

Zwickl, P., Pfeifer, G., Lottspeich, F., Kopp, F., Dahlmann, B. and Baumeister, W., 1990. Electron microscopy and image analysis reveal common principles of organization in two large protein complexes: groEL-Type proteins and proteasomes. *Journal of Structural Biology*, 103(3), pp.197-203.

10. Appendix

	GroEL	GroEL-Rubisco	GroEL-Rubisco	GroEL-Rubisco	GroEL-Rubisco	GroEL-Rubisco + ATP	GroEL-Rubisco + ADP·BeF_x	GroEL-Rubisco + ADP·AlF_x	GroEL-Rubisco + ADP·BeF_x	GroEL-Rubisco + ADP·BeF_x	GroEL-GroES-Rubisco + ADP·AlF_x
Description	Apo data	Untilted data	Tilt data	Chameleon grid 1	Chameleon grid 2	Tilt data	Tilt data	Tilt data	Chameleon grid 1	Chameleon grid 2	Chameleon data
Concentration (µM)	1.5 µM	0.8 µM	55 nM	3.4 µM	3.4 µM	1 µM	1 µM	1 µM	7 µM	7 µM	7 µM
Grid	C-flat 2/2	C-flat 2/2	C-flat 2/2 + GrOx	Self-wicking grid	Self-wicking grid	UltrAufoil 2/2	UltrAufoil 2/2	UltrAufoil 2/2	Self-wicking grid	Self-wicking grid	Self-wicking grid
Freezing instrument	Vitrobot	Vitrobot	Vitrobot	Chameleon	Chameleon	Vitrobot	Vitrobot	Vitrobot	Chameleon	Chameleon	Chameleon
Microscope	Titan Krios	Titan Krios	Titan Krios	Titan Krios	Titan Krios	Titan Krios	Titan Krios	Titan Krios	Titan Krios	Titan Krios	Titan Krios
Acceleration voltage (kV)	300	300	300	300	300	300	300	300	300	300	300
Camera	K2	K2	K2	K2	K2	K3	K3	K3	K3	K3	K3
Collection mode	Counting	Counting	Counting	Counting	Counting	Super-resolution	Super-resolution	Super-resolution	Super-resolution	Super-resolution	Super-resolution
Pixel size (Å)	1.090	1.048	1.048	1.340	1.340	0.524	0.534	0.534	0.534	0.534	0.425
Stage tilt (°)	0	0	30 & 40	0	0	35	-30	-30	0	0	0
# Frames	50	50	40	48	48	50	50	50	50	50	50
Integration time (s)	15	15	8	12	12	3	3	3	3	3	3
Electron exposure (e-/Å)	46.5	57.9	39.3	40.2	40.2	49.1	50.9	51.0	50.6	49.0	72.3
Defocus range (µm)	-1.0 – -4	-1.4 – -3.4	-1.5 – -2.7	-1.4 – -3.0	-1.4 – -3.0	-0.5 – -2.7	-1.5 – -2.7	-1.5 – -2.7	-1.5 – -2.7	-1.5 – -2.7	-1.5 – -2.7
# Micrograph movies	4008	3944	4868	4603	3824	5081	1770	1930	4877	5332	11733

Appendix 1. Data collection parameters for cryoEM data sets.

	GroEL-Rubisco	GroEL-Rubisco-ADP·BeF_x	GroEL-GroES-Rubisco-ADP·AlF_x
Map resolution (Å)	4.5	3.3	3.8
Resolution range (Å)	4.25 – 17.95	3.10 – 7.09	3.57 – 6.54
# Residues	7336	7322	8008
# Non-hydrogen atoms	53914	53872	59426
Ramachandran outliers (%)	0.00	0.00	0.00
C-β outliers (%)	0.00	0.02	0.00
Average B-factor (Å^2)	149.73	126.57	72.16
Clash score	5.03	0.30	5.62
Molprobity score	1.27	1.39	1.77
Map-Model FSC (0.5)	4.4	3.2	3.7

Appendix 2. Model validation statistics. Final atomic models were refined against the locally-sharpened cryoEM reconstructions derived from Chameleon cryoEM data sets. Statistics were calculated by cryoEM validation tools in Phenix and Molprobity.